In van Hout, R., Hulk, A., Kuiken, F. & Towell, R. (2003), *The Lexicon-Syntax Interface in Second Language Acquisition* . Amsterdam: John Benjamins. pp. 151-174.

Inducing abstract linguistic representations: human and connectionist learning of noun classes

John N. Williams
University of Cambridge

Noun class information is a crucial component of the interface between the lexicon and the grammar. In order to explain linguistic productivity it is necessary to assume that linguistic rules are defined not over specific words, but classes of word. This is not only true given the classical distinction between lexicon and grammar, but also in 'emergentist' views which see no clear separation between these two systems (Ellis, 1998; Tomasello, 2000). Even though the latter stress the lexical-specificity of many 'grammatical rules', it is still recognised that adult productivity can only be explained if words are grouped into classes, even if those classes do not map neatly onto traditional linguistic categories. The way in which words are grouped into grammatical classes is therefore an important issue in understanding language development, particularly in explaining the leap from lexical learning to grammar learning.

Noun classes, such as grammatical gender, are fundamentally abstract, grammatical, notions (Corbett, 1991). However, attempts have been made to uncover subtle phonological and semantic cues that can be used to predict a word's gender (Kelly, 1992). For example, masculine nouns in German are more likely to be monosyllabic, and monosyllabic words that are masculine contain more consonants than those of other classes. In French, feminine nouns tend to end in closed stressed syllables (e.g. *personne, tomate, viande*), and masculine nouns tend to end in open stressed syllables (e.g. *avion, bruit, chapeau, bain*). There are also a number of characteristic derivational morphemes associated with each gender (e.g. *-eur* and *-ment* are masculine, and *-tion*, *-euse*, *-iere* are feminine). (Sokolik & Smith, 1992) trained a connectionist network to classify French nouns as either masculine or feminine. The network was presented with the orthographic, rather than phonological, forms of the words. They found that it could then indicate the gender for nouns that it had not received during training, although its performance was

not perfect (ranging between 73% and 75%). This indicates that there are regularities in the form (in this case spelling) of French words which can to a certain extent predict gender category.

Yet there are always words which fall stubbornly outside such generalisations. In the case of French, Carroll (2001) argues that in any case, the kinds of phonological cues that have been appealed to are more subtle than could reasonably be expected to be represented in the lexicon. This is not to say that phonological and semantic cues do not play a role in learning gender systems, or that they do not affect how easy it is to remember the gender of specific words. But ultimately gender classes impose an abstract categorisation on words which is independent of their phonological and semantic properties. Learning gender systems, then, requires the formation of abstract grammatical categories, and producing grammatically well-formed utterances involves applying agreement rules which make reference to those categories.

There is a growing body of evidence which suggests that even quite advanced second language learners continue to make gender errors (Hawkins, 2001; Holmes & de la Batie, 1999). In contrast, such errors are relatively rare in first language acquisition (Caselli, Leonard, Volterra, & Campagnoli, 1993). There is also evidence for qualitative differences between first and second language acquisition and processing of gender. A number of studies have shown that second language learners are more sensitive to phonological agreement patterns that correlate with gender classes than either children or adults in their native language. For example, for the Italian *il pettine* (the comb, masculine singular) a second language learner might produce **le pettine*, using the article which is more often associated with the *-e* ending on feminine plural nouns (Holmes & de la Batie, 1999). In contrast a child would be more likely to produce **il pettino*, choosing an article that is correct for the noun's gender and number, and providing the noun with the characteristic *-o* ending for masculine singulars. This demonstrates a grasp of the noun's abstract gender as the controlling influence in determiner selection (Caselli et al., 1993). In reaction time tasks on adults, Taraban & Kempe, (1999) showed that non-native speakers of Russian are more sensitive to phonological cues to gender than are natives. Finally, a study by Guillelmon & Grosjean (2001) showed that whereas native speakers of French and early bilinguals show certain gender congruency effects in reaction time

tasks, such effects are absent in late bilinguals. These studies suggest that second language learners do not achieve native-like representation or processing of gender information.

In this chapter I shall explore the possibility that the reason why gender is a persistent problem for second language learners is precisely because the underlying abstract grammatical concepts are difficult to acquire through associative learning. I shall address this issue through behavioural studies of semi-artificial language learning in tandem with computational (connectionist) simulations. These simulations were used as a means of assessing the viability, and potential limitations, of a purely associative learning account of the behavioural data.

The issue of abstraction in human and connectionist learning

Noun class induction provides a well-constrained domain in which to examine the broader issue of abstraction in both human and connectionist learning. In the case of adult implicit learning there has been a good deal of debate over whether the knowledge that is acquired in, say artificial grammar learning experiments can really be characterised as abstract (compare Johnstone & Shanks, 1999; Knowlton & Squire, 1996; Meulemans & Van der Linden, 1997). Some degree of abstraction is suggested by the ability to transfer rule knowledge between stimulus sets (Knowlton & Squire, 1996; Mathews et al., 1989). But this appears to be no more than knowledge of patterns of alternation or doubling of stimuli, for example the common abstract ABA structure which underlies the syllable sequences *ga-ti-ga* and *wo-fe-wo* (Marcus, Vijayan, Bandi Rao, & Vishton, 1999). Gómez & Gerken (2000) refer to this as “pattern-based abstraction”. But language structure depends upon patterns that are defined over abstract categories, such as the common NVN structure underlying 'Dogs eat pizza' and 'John loves books'. Gómez & Gerken, (2000) refer to this as "category-based abstraction". Very little implicit learning research has examined this kind of abstraction, even though it is a prime area in which implicit learning of language structure can be evaluated.

In connectionist networks rule-like behaviour, such as the ability to generalise to novel inputs, is an emergent property of the system, and there is no separation between

rote memory for examples and the representation of underlying generalisations (consider, for example, the well-known models of past tense formation Rumelhart & McClelland, 1986, and reading, Seidenberg & McClelland, 1989). But it has been argued that the human generative capacity in linguistic domains can not be accounted for without the 'classical' distinction between knowledge of instances and knowledge of rules, or the traditional computational distinction between data and symbolic programs (Fodor & Pylyshyn, 1988). According to this view, the problem with connectionist models is that they respond to novel inputs purely on the basis of their similarity to trained examples, and not by applying abstract rules (Berent, Marcus, Shimron, & Gafos, 2002; Marcus, 1999; Marcus et al., 1999). Category-based abstraction provides an ideal arena in which to explore this issue.

Previous research into human and connectionist learning of word classes

In his work on sequence learning Elman (1990) showed that there is a sense in which a connectionist network can learn abstract noun classes. This network learned the sequential probabilities of words in simple sentences through a prediction task (attempting to predict the next word in a sentence on the basis of the preceding ones). When the internal states of the network were examined (see below for an illustration of how this is done) it was found that the activation patterns produced by words clustered into classes that reflected the distributional properties of the training sentences. The two largest clusters were for nouns versus verbs, and within these groups there were smaller sub-clusters corresponding to transitivity preference for verbs, and animacy for nouns. These clusters were based purely on a distributional analysis of the words in the input. For example, what made a noun 'inanimate' was nothing more than the fact that it only occurred before certain kinds of verb (e.g. *move*, *break*) and not others (e.g. *smell*, *see*). This work is widely cited as proof that networks can induce word classes by performing distributional analysis, and as support for a statistical approach to language learning (Redington & Chater, 1998).

Given the apparent power of distributional information to deliver noun class information it is perhaps surprising that there is only limited evidence from experimental

studies that humans are able to exploit it in order to learn noun classes. Saffran (2001) examined incidental learning of a set of hierarchical phrase structure rules in which each phrase was associated with a distinct class of nonsense words. She argued that the results of the grammaticality judgement tests showed that the participants developed sensitivity to phrase structure and word class, and that this was based on a statistical analysis of the distribution of the words in the input. However, abstract representations of word class would permit test items containing word sequences that had never occurred in the input to be judged as grammatical (or more grammatical than similar sequences which violated phrase structure). Because no such test was performed it is difficult to know whether abstract word classes had really been learned.

More stringent tests of word class learning become possible when noun classes, such as gender systems, are considered. Brooks, Braine, Catalano, & Brody (1993) used an artificial language in which there were two noun classes, and each class used different affixes to mark the location of the actor in relation to the object denoted by the noun. Neither the form nor meaning of the nouns provided any clue to their class. Adults were first taught the vocabulary, and then performed both comprehension and production tasks (e.g. acting out phrases, or describing pictures with feedback in the form of the correct answer). After training they were tested on knowledge of the trained items, and also on their ability to produce the correct response for noun-affix combinations that had not been presented during training. Whilst their performance on trained items was at around 75%, they were at chance on the generalisation items. Not one of the 16 subjects showed evidence of having learned the system. Similar results have been obtained in a number of other studies (Braine, 1987; Braine et al., 1990; Frigo & McDonald, 1998). Frigo & McDonald (1998) argue that models of noun class learning that depend on pure distributional analysis (Anderson, 1983; Maratsos & Chalkley, 1980; Pinker, 1984) are "too powerful" (ibid. p. 237).

The question is, then, does connectionism fall into this class of overly powerful learning mechanisms for learning noun classes? The experiments and simulations presented below further explored the circumstances under which arbitrary and non-arbitrary noun class systems can be learned by humans and connectionist networks.

Experiment 1

Williams & Lovatt (in press) tested whether humans can learn the arbitrary noun class system shown in Table 1. There were eight nouns divided into two arbitrary classes 'masculine' and 'feminine'. Words in the 'masculine' class occurred with the determiners *ig*, *i*, *ul*, and *tei*. Words in the 'feminine' class occurred with the determiners *ga*, *ge*, *ula*, and *tegge*.[1] The training items were the non-italicised phrases shown in Table 1. The italicised items were withheld for testing generalisation. It would only be possible to know that 'the ball' should be translated as *ig johombe* by knowing that *johombe* belongs to the 'masculine' class. Neither its form, its *–e* ending, nor its meaning provide any clues.

Table 1: *The items employed in Experiments 1 and 2. Items used for testing generalisation are in italics.*

	definite singular (the)	definite plural (the)	indefinite singular (a)	indefinite plural (some)
"masculine"				
ball	*ig johombe*	i johombi	ul johombe	tei johombi
house	ig zabide	*i zabidi*	ul zabide	tei zabidi
fight	ig wakime	i wakimi	*ul wakime*	tei wakimi
bird	ig migene	i migeni	ul migene	*tei migeni*
"feminine"				
shoe	*ga shosane*	ge shosani	ula shosane	tegge shosani
kiss	ga tisseke	*ge tisseki*	ula tisseke	tegge tisseki
cake	ga chakume	ge chakumi	*ula chakume*	tegge chakumi
nose	ga nawase	ge nawasi	ula nawase	*tegge nawasi*

The participants first learned the nouns and determiners as isolated vocabulary items. They then received the determiner-noun combinations for each training item as part of an exercise in rote memorisation that cycled through phases of presentation and cued recall over sets of 4 items. Four phrases were presented with their English translations, for example: 'the nose' - *ga nawase*, 'the birds' - *i migeni*, 'some balls' - *tei*

johombi, 'a kiss' - *ula tisseke*. The participants repeated each novel phrase immediately after they had seen and heard it. After the four phrases had been presented participants attempted to recall each phrase given the English translation and stem as cues, for example: 'the birds' _ *migen*_ , 'the nose' _ *nawas*_ , 'a kiss' _ *tissek*_ , 'some balls' _ *johomb*_ . They were provided with feedback after each recall attempt in the form of the correct answer. After receiving the 24 training items they performed a generalisation test on the withheld items in Table 1. The generalisation test was similar to the recall component of the training phase. The English translation of each phrase was presented (e.g. 'the ball'), along with the form of the corresponding stem (*johomb*_), and the participants had to produce the appropriate determiner and appropriately inflected noun. No feedback was given. This sequence of memory and generalisation tasks was repeated five times.

Across 21 participants the mean generalisation performance over the five cycles was 36%, 48%, 54%, 66%, and 67%. A repeated measures ANOVA showed that the improvement in performance was significant, $F(4,80) = 13.11$, $p<0.001$. This shows that the participants learned something of the underlying noun class organisation. However, there were large individual differences in the level of learning. Two factors were found to independently predict performance on the final generalisation test. The first was the participants' phonological short-term memory, as measured prior to the experiment by their ability to recall lists of 3 nonsense words (the singular forms of the nouns in the target language) in the order of presentation. The correlation between this memory measure and performance on the final generalisation test was $r = 0.528$, $p<0.05$. There was evidence that the relationship between phonological short-term memory and rule learning was mediated by memory for determiner-noun combinations received during training. Clearly memory ability is crucial to performing the kind of distributional analysis upon which learning of this kind of system depends.

The second factor was a measure of the participants' breadth and depth of knowledge of other gender languages. All of the participants' L1s were non-gender languages (in fact all but one of them was a native speaker of English), but the more gender languages they knew as L2s, and the better they knew them, then the better their

performance on the generalisation test ($r = 0.520$, $p<0.05$). This suggests that the learning process was facilitated by linguistic knowledge.

There are a number of possible reasons why our participants managed to learn an arbitrary noun class system whereas those in the previous studies did not. First, the systems used by Brooks et al. (1993) and Braine et al. (1990) involved agreement between spatial prepositions and nouns, and Frigo & McDonald (1998) used a system involving agreement between greetings and names. Participants may have had relatively little familiarity with similar systems in other languages that they knew. Second, it is possible that the size of the languages is important. Braine et al. (1990) used a 24-word vocabulary, Brooks et al. (1993) used 30 words, and Frigo & McDonald (1998) used 20 words, whereas Experiment 1 used only 8 words. Clearly, keeping track of the collocates of 20 to 30 words is much harder than keeping track of the collocates of 8 words. A third potentially important factor is that in the present case some of the determiners in each class had the same ending. The feminine class contained the pairs *ga-ula* and *ge-tegge*; the masculine class contained *i-tei*, and the remaining determiners *ig* and-*ul* were the only ones to end in consonants. This similarity structure may have facilitated the learning process.

Experiment 1 demonstrates that an arbitrary noun class system is in principle learnable. The question now is whether a connectionist simulation of the same learning problem will be similarly successful.

Simulation 1

For this and all other simulations reported here, the simulation package *Tlearn* was used (Plunkett & Elman, 1997). The aim in the first simulation was to train the network in a way which resembled as closely as possible the training task performed by the participants in Experiment 1. I decided to focus on the recall component of the training task. The network was taught to produce the correct determiner for each phrase in the training set shown in Table 1. The input consisted of representations of the noun stem, the inflection, the English determiner, and the number of the noun. For example, the input for the item *tei johombi* was ‘johomb’, ‘-i’, ‘some', and ‘plural’. This is the

information that is relevant to predicting the determiner, and which was explicitly provided to the participants in the recall component of the training task in Experiment 1.[2] Following Elman (1990) one unique input node was used to represent each element of the input (for example one unit was used to represent *johomb*), yielding a total of 15 input nodes (8 stems, 2 inflections, 3 English determiners, singular, plural). The input nodes were connected to 5 hidden units, which were in turn connected to 8 output units, one for each of the 8 possible determiners. For each input pattern the network was taught to produce the correct determiner. For example given the input *johomb*, *-i*, 'some', and 'plural' it was taught to predict *tei*. This involved comparing the actual output from the network with the correct output, and making appropriate changes to the connection weights within the network according to the degree of error. In this sense the network was provided feedback in the same way as the participants in the experiment.

The network was initially trained until the root mean square (RMS) error for the *training* items was 0.1 (this required an average of 2,479 cycles through the training set).[3] An error of this magnitude indicated that for each input pattern the network was able to activate the correct determiner on the output layer to a value close to the target value of 1.0, and all other output units had values close to zero. Testing involved presenting the input patterns for the *generalisation* items in Table 1 (i.e. the network was presented with patterns that it had not received during training). For each input pattern, the activation level of the output units was recorded, compared to the correct answer and the degree of error calculated. The training and test procedure was repeated 20 times, and on each run the connection weights were given random starting values.

Generalisation performance on each run was perfect in the sense that the activation on the node for the correct determiner was far greater than that of the others. Over 20 runs the mean RMS error was 0.118 (which is not much greater than that for trained items). That is, the network was able to correctly predict the determiner for input patterns that it had never encountered during training with an accuracy which was almost as high as for the trained items.[4]

In order to explore the nature of the network's internal representations the word stems were presented alone to the input layer at test (i.e. all of the other elements of the input were given values of zero). The activation patterns over the 5 hidden units were

recorded and submitted to a cluster analysis (for a similar procedure see Elman, 1990). The logic of only presenting the word stems was that the aim was to ascertain the similarity structure of the hidden unit activations to the nouns in a way that was not contaminated by the activations produced in specific contexts of definiteness and number. Over 6 separate runs a similar result was obtained -- the activation patterns clustered according to gender. That is, nouns within the same class produced activation patterns that were similar to each other and distinct from the patterns produced by the nouns in the other class.

It should be clear that this network is not simply producing responses to the generalisation items on the basis of their similarity to trained items. For example, for the test item *ig johombe* the stimuli were *johomb*, *-e*, 'the', and 'singular'. During training *johomb* and *-e* only occurred with *ul*. The elements *the* and 'singular' occurred with both *ig* and *ga* with equal frequency. Yet the network was able to produce a strong output on *ig* and much lower levels of activation on the remaining determiners. Simulation 1 therefore shows that a connectionist network can achieve linguistic productivity, and can behave *as if* it has formed abstract representations, even though there are no abstract representations as such within the network.

There are various ways in which the power of Simulation 1 could be varied in order to account for the effects of individual differences in Experiment 1, or the failures to obtain learning of arbitrary noun classes in previous experiments. The effect of memory ability could be dealt with by changing the learning rate parameter (which determines the size of the weight changes in response to a given amount of error). Factors such as the similarity structure of the determiners, or the number of nouns in the training set, would be expected to influence learning rate as well. However, the influence of knowledge of other gender languages is more problematic and will be considered after the remaining experiments and simulations have been reported.

Connectionist networks are commonly regarded as models of the associative mechanisms underlying implicit learning (Cleeremans & Jiménez, 2002). However, when we debriefed our participants after Experiment 1 it was clear that the more successful amongst them had been employing intentional learning strategies, and that there was a good correspondence between their conscious understanding of the system and their

performance in the final generalisation test. It therefore becomes important to test whether learning could be obtained under implicit conditions.

Experiment 2

This experiment employed a training task that was thought to be unlikely to induce an intentional learning strategy. Participants first performed the same phonological short-term memory test and vocabulary learning exercise as in Experiment 1. Determiner-noun combinations from the training set were then auditorily presented in a semi-random sequence, avoiding immediate repetitions of the same noun or determiner. For each item the participants had to perform the following tasks: (I) repeat the phrase aloud, (II) indicate whether it refers to a living or non-living thing by pressing one of two response keys, and (III) translate the phrase into English. For example, for the item *ul johombe* they would respond by saying “ul johombe”, pressing the non-living key, and saying "a ball". The meanings of the words were altered so that half of the nouns in each class referred to living things and half to non-living things. The living/non-living decision was included because this experiment was also a control for a subsequent version in which noun animacy predicted noun class membership (see Experiment 3 below). Here it serves as a means of increasing task demands so that participants would be less likely to attempt to engage explicit learning processes. The participants were told that the purpose of the experiment was to see how their decision and translation performance improved with practice and so they were encouraged to make their responses as quickly and as accurately as possible. Training extended over 15 cycles through the 24-item training set, giving a total of 360 training trials. This took between 60 and 75 minutes including rest breaks after every 5 cycles.

The training phase was followed by the test phase. On each trial the English translation of a test phrase was visually presented (e.g. ‘the ball') and the participants had to choose between a grammatical and ungrammatical translation in the target language, where the determiner for the ungrammatical item was always of the correct number and definiteness, but the incorrect gender (e.g. *ig johombe* versus *ga johombe* for 'the ball').

First the 8 generalisation items were presented (see Table 2) followed by 16 trained items.

There were 18 participants who were selected on the basis of their good knowledge of gender languages so as to increase the potential for obtaining learning in this experiment. They all rated themselves as intermediate or better in at least two gender languages (mean = 2.8, range = 2 to 6). Twelve of the participants spoke a gender L1. Using the same scale for assessing knowledge of gender languages as employed by Williams & Lovatt (in press) they scored 5.8, which is much higher than the mean of 2.6 for the participants in Experiment 1. Their phonological short-term memory was also somewhat superior, the mean score being 71% as opposed to 64%.

None of the participants were aware of the noun class system either during training or test phases. The average percentage correct on the generalisation items was 56%, which was not significantly different from the chance level of 50%, $t = 1.34$, $p > 0.1$. On the other hand, performance on the trained items was 69%, which is significantly better than chance, $t = 5.53$, $p < 0.001$, and significantly better than performance on generalisation items, $t = 2.58$, $p < 0.05$. Thus, although the participants had quite good memory for trained items, there was no evidence of learning the underlying noun class distinction. This conclusion is emphasised by the fact that the 10 participants who scored 75% or better on the trained items (mean = 80%) had a mean generalisation score of 50%. Nor were there any correlations between generalisation test performance and either phonological short-term memory or language background, and participants who spoke a gender L1 did no better than those than did not (generalisation test scores were 56% for both groups).

Given the failure to obtain learning in this experiment one may conclude that Simulation 1 was in fact too powerful, and that the learning that occurred in Experiment 1 was a result of purely explicit processes which fall outside the scope of the model. However, there is an alternative possibility. We should also consider the relationship between the task performed by Simulation 1 and the tasks performed by the participants in Experiments 1 and 2. Simulation 1 was intended as a model of the recall component of the training task used in Experiment 1. But in Experiment 2 the participants' task was very different. They did not have to generate any determiners at any point during

training, but only had to perform animacy decisions and produce English translations. Simulation 1 could not be said to be a good model of this task. A second simulation was therefore conducted that made different assumptions about the learning task.

Simulation 2

Incidental learning is best regarded as a relatively passive process of recording correlations between attended features in each experience. Cleeremans & Jiménez (2002), following O'Reilly & Munakata (2000), have referred to this as 'model learning', the goal of which is to "enable the cognitive system to develop useful, informative models of the world by capturing its correlational structure" (*ibid*. p. 18). Connectionist models of model learning do not require feedback because the system merely attempts to represent the structure of the inputs it is provided. This is in contrast to 'task learning' which has the aim of "mastering specific input-output mappings (i.e. achieving specific goals) in the context of specific tasks through error-correcting learning procedures" (*ibid*. p. 18). Crucially for present purposes they assume that model learning operates continuously, regardless of the task. Simulation 1 instantiated task learning, and was successful because the underlying noun class distinction happened to be relevant to the task the network was required to perform. But in Experiment 2 the tasks that the participants were performing (animacy decisions and translation) exerted no pressure to learn the noun class distinction. The same would be true of simulations of those tasks. The only way in which the noun class distinction could be learned, therefore, would be through model learning, which requires a different kind of network from that used in Simulation 1.

One way of instantiating model learning is to train a three-layer network to associate each input *to itself*. That is, the network learns to reproduce the input pattern on the output layer. These are called "autoassociation" networks (Plunkett & Elman, 1997). Because there are fewer hidden than input/output units the network is forced to discover an economical means of representing the patterns so that they can be reproduced on the output. This gives the network the potential to extract generalisations. Autoassociation networks do not require feedback because the input itself provides the reference point

against which the accuracy of the output can be judged. How does such a network fare on the arbitrary noun class induction problem?

In Simulation 2 there were 31 input units representing the 8 determiners, 8 stems, 2 inflections, 3 English determiners, 8 English nouns, and units for singular and plural. All of the relevant information in a training item such as *'ul johombe*, a ball' was represented as a pattern over the input layer. The 31 output units represented the same information as the input units. The network had 20 hidden units.[5] For each item in the training set the network was trained to reproduce the input pattern on the output layer. Training continued until output error ceased to decline (which was after about 2,500 cycles).

In Experiment 2, learning was assessed by forcing participants to choose between two translations for a phrase, for example, between 'ga johombe' and 'ig johombe' as translations of 'the ball'. The model can be tested in the same way by presenting both grammatical and ungrammatical determiner-noun combinations and comparing the strength of the output on the determiner units. For a trained item, such as *ul johombe*, the strength of activation of the corresponding determiner in the output, in this case *ul*, was, as one would expect, very high (0.996 when averaged over 8 training items on 5 separate runs, where the required activation level was 1.0). Ungrammatical items such as *ula johombe* produced much weaker activation of the corresponding output determiner node, in this case *ula* (0.214). Clearly the network had not simply learned to reproduce input patterns on the output layer. Rather, its ability to do so was affected by whether it had received those patterns during training. In human terms this would be the equivalent of a greater feeling of familiarity for *ul johombe* than *ula johombe*. But for generalisation items the output activation on determiners in both grammatical and ungrammatical items, e.g. *ig johombe* versus *ga johombe*, was very low and not significantly different (0.054 and 0.055 respectively). In other words, both items appeared equally unfamiliar to the network. Therefore, like the human participants in Experiment 2, the autoassociation network had good memory for trained items, but was unable to distinguish between grammatical and ungrammatical generalisation items.

The contrast between Simulations 1 and 2 demonstrates that task learning enabled a connectionist network to become sensitive to an abstract noun class distinction whereas

model learning did not. This is a rather surprising result when one considers that there is a sense in which the networks were performing rather similar tasks. In both cases they had to remember which determiners occurred with which configurations of noun, definiteness, and number in the training items. The difference was that in Simulation 1 the network's resources were focused on predicting the determiner from the cues that it was provided, whereas in Simulation 2 the network was actually attempting to remember the unique combination of determiner, noun, definiteness, and number that occurred in each training item. This exercise in episodic memory for entire training episodes apparently did not exert sufficient pressure on the network to discover the underlying noun class distinction.

The contrast between task learning and model learning is reminiscent of the procedural-declarative distinction in Anderson's ACT framework (Anderson, 1983). Productions are sets of rules which match their 'IF' conditions against the current contents of working memory, and if these are satisfied, they 'THEN' produce some action, or deposit some other kind of representation in working memory. Although stated in a symbolic formalism in ACT, a connectionist network can be conceived as a subsymbolic model of the entire set of productions which perform the transformation between one type of input to another type of output (Sun et al., 2001). Both procedural learning and connectionist learning of this type depends on error correction. In contrast, Simulation 2 could be identified with the 'declarative' memory component of the ACT framework. The idea that these two kinds of memory system might have differential power to extract generalisations from the environment is clearly relevant to attempts to construct a theory of second language acquisition in terms of their interaction (Towell & Hawkins, 1994), or to identify them with different brain regions (Ullman, 2001). Indeed, the idea that procedural learning is more powerful at extracting abstract linguistic rules would be consistent with the proposal that such a mechanism supports first language acquisition, whereas second language acquisition is supported by declarative learning (Ullman, 2001).

If Simulation 1 is accepted as a valid model of the learning process in Experiment 1 then there is another interesting consequence. The learning that was occurring in that experiment was characterised as 'explicit'. Not only did the participants appear to have an intention to learn, but some of them also made comparisons between consciously recalled

input items, and formed conscious hypotheses. Simulation 1 captures the intentional component of the learning process, since it too evaluated its outputs with respect to feedback for the purpose of learning in order to be able to generate determiners. But it obviously does not model the other components of what, in human terms, we regard as explicit learning. However, this does not necessarily detract from the relevance of the model. Shanks (1995) reviews a range of studies on human learning where there is a good fit between human behaviour and connectionist models. Yet in many of these experiments the participants were actively searching for rules. For example, in a medical diagnosis task (see Shanks, 1995, p. 42) participants were presented with hypothetical patients with certain symptoms and were instructed to diagnose what illness each patient had. Each trial was accompanied by feedback in the form of the correct diagnosis. Performance was directly related to the degree of contingency between different cues (symptoms) and outcomes (diseases) in the training data. Shanks showed that the results could be adequately modelled by a simple connectionist network in which symptoms were presented as inputs, diagnoses as outputs, and the correct diagnosis was provided as feedback (Shanks, 1995, p.120). Yet the participants in the experiment presumably had the experience of actively trying to work out the relationship between symptoms and diseases. Whilst it is presently unclear how the conscious states of the learner influence the learning mechanism, it should not be assumed that the possibility of there being such interactions rules out a unified associative explanation (Cleeremans & Jiménez, 2002).

Experiment 3

With the benefit of the hindsight Experiment 2 was a poor test of implicit learning because the kind of associative learning mechanism supposed to underlie implicit and incidental learning would not be expected to learn the underlying generalisations. We[6] therefore decided to run Experiment 2 again, but this time using a system that would be learnable even by the kind of autoassociation network used in Simulation 2. The language was essentially the same as that shown in Table 1 except that the meanings of the words were altered so that all of the words in Class I referred to living things and all of the words in Class II referred to inanimate objects. For simplicity, the living/non-living

distinction will be referred to here in terms of an 'animacy' cue to noun class. It has been shown that under intentional learning conditions humans have no problem grasping semantically-based noun classes (Braine, 1987; Carroll, 1999). A version of Simulation 2 that included animacy information confirmed that the present system was also learnable by an autoassociation network. This is presumably because there are direct associations between the units that encode animacy and certain determiners. Note, therefore, that in this experiment we are no longer concerned with whether implicit learning of abstract noun classes is possible. Rather the issue is whether implicit learning of a noun class distinction can be obtained under conditions where the connectionist model predicts that there should be an effect.

The tasks and procedure were exactly the same as in Experiment 2. For each phrase presented during training the participants had to repeat it, indicate whether it referred to a living or nonliving thing, and translate it into English. Note that this time the living/nonliving decision coincided with the noun class of the word. The same learning tests were used as in Experiment 2. There were 37 participants with varied language backgrounds.

Only seven of the participants became aware of the noun class distinction and its relation to animacy during the training phase, and their performance was perfect, or near perfect, on the generalisation and trained items. None of the remaining 30 participants became aware of the system during the training phase and none of them claimed to have been consciously trying to work out the system during the generalisation test. Even at the end of the whole testing phase none of them realised the relevance of animacy. Nevertheless, performance on generalisation items was 61%, which was significantly above the chance level of 50%, $t = 3.25$, $p < 0.01$. They scored 71% correct on trained items, which was also significantly above chance, $t = 6.09$, $p < 0.001$. Therefore, Experiment 3 succeeded in demonstrating at least some degree of implicit learning of a system that was also learnable by an autoassociation network.

However, there were large individual differences in generalisation test performance. Just as in Experiment 1 there were correlations with phonological short-term memory ($r = 0.50$, $p < 0.01$) and knowledge of gender languages ($r = 0.586$, $p < 0.001$), which in this case was quantified simply in terms of the number of gender

languages in which the participants rated their proficiency as intermediate or better (mean = 1.8, range = 0 to 5). We also evaluated whether, amongst the 30 unaware participants, speakers of gender L1s did better than speakers of non-gender L1s. For the 13 speakers of gender L1s mean generalisation test performance was 71%, which is significantly above chance, $t = 4.08$, $p < 0.01$, whereas for the 17 speakers of non-gender L1s it was 54%, which is not significantly above chance, $t = 0.96$. The difference between these two groups was significant, $t = 2.78$, $p < 0.01$. The two groups did not differ significantly in terms of the number of L2s spoken to an intermediate level or better (3.54 and 3.12 respectively, $t < 0.92$), the number of gender languages known as an L2 (the means were 1.46 and 1.23 respectively), but they did differ slightly in terms of phonological short term memory (77% versus 68%, $p = 0.08$). Better matched groups resulted from removing the three participants with the lowest phonological short term memory scores from the sample (all scores were less than 50%, and all three participants were in the non-gender L1 group). The 13 speakers of gender L1s and remaining 14 speakers of non-gender L1s were well matched in terms of number of gender languages spoken as an L2 (1.46 and 1.43 respectively) and in terms of phonological short term memory scores (77% and 73%). Yet the generalisation scores were 71% and 55% (the difference being significant, $t = 2.31$, $p < 0.05$). Note that for the non-gender L1 group the mean for the trained test items was well above chance (67%, $t = 3.74$, $p < 0.01$).

Discussion

In one sense it could be argued that there is a good alignment between the connectionist models and the human data in the present studies, provided assumptions are made about which kind of network is appropriate to which task conditions. Where the model was able to generalise there was also evidence for generalisation amongst the participants in the experiment (Simulation 1 and Experiment 1, Simulation 2 supplemented by animacy information and Experiment 3). Where the model was not able to generalise there was no evidence for generalisation amongst the human participants (Simulation 2 and Experiment 2). The problem is, however, that the networks only seem

to account for learning amongst those participants who already possessed knowledge of other gender languages. Yet none of the networks contained any prior knowledge. Seen in this light they provide a poor fit to the human data. In this final section I shall consider ways in which prior knowledge could have influenced human learning, and whether the data then become more amenable to a connectionist interpretation. I shall then consider the implications of the present results for second language acquisition.

The role of prior linguistic knowledge

One way in which prior knowledge could facilitate learning is through its effect on the learners' strategy. Recall that the success of Simulation 1 depended upon using number, definiteness, and an abstract representation of the nouns (represented as single nodes) to generate the determiners. But this presupposes a certain understanding of the nature of gender systems. Participants who did not have this understanding may simply have approached the task in the way that it was presented to them; that is, as a short-term memory exercise for determiner-noun combinations. In that case their learning processes would be more appropriately modelled by Simulation 2 than Simulation 1. Indeed, the contrast between Simulations 1 and 2, between task learning and model learning, could be seen as a computational account of a more general contrast between analytic and non-analytic, memory-based, learning strategies (Skehan, 1998). In the present case the probability of adopting an *appropriate* analysis strategy could have also depended upon metalinguistic knowledge of other gender systems that was derived from second language learning experience.

Obviously a learning strategy account can not apply to the kind of incidental and implicit learning occurring in Experiment 3. However, in this case learning failures could be accounted for simply by assuming that animacy was not perceived as being relevant to the determiners. In Williams (in preparation) I argue that implicit learning of form-meaning connections (such as between determiners and animacy information) is problematic because of the requirement that form and meaning are unitized at encoding; learners must actually perceive them as being relevant to each other. Merely paying attention to the relevant elements does not appear to be sufficient, at least not under the

task conditions of Experiment 3. In terms of the model learning mechanism instantiated by Simulation 2 this means that even though animacy information was attended, it did not enter the same memory trace as information about the determiner, noun identity, definiteness, and number. The problem is, therefore, to explain why participants who spoke a gender L1 defied this principle and were able to unconsciously associate the determiners with animacy information. There is no obvious connectionist answer to this problem. Is the classical linguistic approach any more promising?

Linguistic (Carroll, 1989; Hawkins, 2001) and psychological (Levelt, Roelofs, & Meyer, 1999; Vigliocco, Antonini, & Garrett, 1997) analyses of gender representation and processing in the L1 assume abstract gender features that are attached to nouns in the lexicon. How gender features are acquired is not often considered. However, Carroll, (2001) proposes an induction procedure which is triggered by the presence of alternating determiner forms in the input (e.g. two words for 'some'). The first occurrence of one of the determiners, for example in *tei johombi* ('some monkeys') has no effect. But when another phrase involving a word for 'some' is encountered, for example *tegge nawasi* ('some vases'), the learner seeks to rationalise the contrast by marking the noun with a [+Gender] feature. In this way, one of the determiners becomes an assigner of the gender feature whilst the other remains the default. Remembering which of the alternating pair of determiners assigns the gender feature is likely to be problematic, however. In the (admittedly artificial) case that [+Gender] also corresponds to some other active feature of the noun, such as [+inanimate], one can imagine that this problem would be alleviated. To account for the influence of gender L1s in Experiment 3 it would have to be assumed that this kind of induction mechanism can only operate in L2 if it was used in the L1. This is perhaps not too implausible if one considers that each time a speaker of a gender language encounters a novel noun the same process of using the accompanying determiner to assign gender to it must operate. On the other hand, it is another matter to assume that, when confronted with a new language, learners are able to assign new gender features on the basis of newly observed alternations between determiners. It is also relevant to consider that at present there is no evidence that speakers of gender L1s have any less problems with gender in an L2 than do speakers of non-gender L1s (Bruhn & White, 2000). Thus, although the gender L1 advantage found in Experiment 3 is

intriguing, there is no obvious way of accounting for it at the present time from either connectionist or classical perspectives.

Implications for second language acquisition

When considering second language acquisition, particularly under naturalistic un-instructed conditions, it is relevant to consider the power of incidental learning mechanisms; that is, learning that takes place as a natural consequence of processing the relevant stimuli for purposes other than discovering the underlying regularities. This means that we should consider implicit learning conditions like those in Experiments 2 and 3 and model learning mechanisms of the type exemplified by Simulation 2 as being the most relevant. Granted this assumption, then the prospects for associative learning of abstract noun classes would appear to be bleak.

However, one limitation of Experiment 2 and Simulation 2 is that they employed a completely arbitrary noun class system. As mentioned earlier, it has been argued that in many natural languages at least a proportion of the members of the same noun class share phonological and semantic properties. Could the presence of these cues facilitate learning? In fact a number of experimental studies have shown that partial phonological and semantic cues do indeed facilitate noun class induction (Braine, 1987; Brooks et al., 1993; Frigo & McDonald, 1998). However, these studies have only demonstrated an effect of partial cues under intentional learning conditions similar to those in Experiment 1. There have been no demonstrations of their effect upon implicit learning. Indeed, my own preliminary investigations of learning such systems using networks of the type used in Simulation 2 have failed to generalise to unmarked words (whereas a network such as Simulation 1 would clearly have no problem).

Even under the intentional learning conditions of the earlier experiments there was very little evidence of generalisation to items that did not carry the appropriate cues. The adults in Brooks et al.'s (1993) study showed barely a significant effect using a one-tailed test (which assumes that the direction of the difference is predicted), and for the children in their second experiment there was no evidence of generalisation at all. Given that 7 out of the 16 adults had explicit knowledge of the word classes, whereas only 1 of

the children did, then it seems likely that these participants were responsible for the slightly above-chance performance of the group as a whole. Generalisation to unmarked nouns would therefore seem to be unlikely under implicit conditions. Only in one of Frigo & McDonald's (1998) three experiments was performance on unmarked generalisation items significantly above chance, and this was when word class was indicated by a characteristic initial and final syllable (e.g. *wan*ersum*glot*, *wan*olov*glot*, *wan*al*glot* versus *kai*salm*rish*, *kai*silvr*ish*, *kai*salb*rish*). Braine (1987) also obtained good generalisation to unmarked words, but half of the nouns in one class referred to males and the other half to females. Thus, generalisation appears to be limited to cases where the cues are more salient than in natural languages.

Somewhat counterintuitively, where the above studies did find evidence of generalisation to unmarked items was when entirely novel nouns were introduced in the final test phase. The equivalent test in the context of the language used here would involve telling participants that *ul vark* means 'a dog' and asking them to produce the translation of 'the dog' (the correct answer being *ig vark*). Such a test only requires knowledge of the associations between the determiners. Therefore, it does appear that partial phonological cues can facilitate acquisition of inter-determiner associations (or rather, their equivalent in the languages that were used). Determiners in the same class presumably become associated by virtue of their frequent association to the same phonological cue. Generalisation is then achieved by a process of *inference* from another determiner-noun combination that is provided at test or recalled from memory. As argued by Frigo & McDonald (1998), poor performance on generalisation tests involving nouns that occurred in training could be because of problems recalling an example of a determiner that occurred with that noun. But the native speaker of a gender language is assumed to generate an appropriate determiner directly on the basis of an abstract specification of the noun's gender in the lexicon, not by inference. It is far from clear that the participants in these experiments acquired knowledge of noun classes in that sense.

The results from these studies do not, therefore, offer much prospect of incidental learning of noun classes. This is of course consistent with the claim that gender is a persistent problem for second language learners. Assuming an underlying model learning mechanism such as that in Simulation 2, learning would be predicted to be limited to rote

storage of determiner-noun combinations, and associations between determiners and partial phonological and semantic cues. This would explain L2 learners' sensitivity to phonological cues in gender processing tasks (Guillelmon & Grosjean, 2001; Holmes & de la Batie, 1999; Taraban & Kempe, 1999). Unmarked nouns would have to be dealt with through rote storage, putting a strain on phonological memory (Williams & Lovatt, 2003). The lack of a true underlying noun class organisation (of the type exemplified in Figure 1) would make storage of determiner-noun combinations particularly prone to error, but if at least one instance of a determiner-noun pair can be retrieved, other appropriate determiners could be inferred using knowledge of inter-determiner associations. Thus, second language learners can acquire a semblance of competence but the failure to organise the underlying representations in terms of abstract noun classes will cause persistent problems. I have argued that this reflects a weakness in the type of associative learning mechanism that is assumed to underlie incidental learning.

References

Anderson, J. R. (1983). *The Architecture of Cognition*. Cambridge MA: Harvard University Press.

Berent, I., Marcus, G. F., Shimron, J., & Gafos, A. I. (2002). The scope of lingusitic generalizations: evidence from Hebrew word formation. *Cognition, 83*, 113-139.

Braine, M. D. S. (1987). What is learned in acquiring word classes: a step towards an acquisition theory. In B. MacWhinney (Ed.), *Mechanisms of language Acquisition* (pp. 65-87). Hillsdale, New jersey: Lawrence Erlbaum Associates.

Braine, M. D. S., Brody, R. E., Brooks, P. D., Sudhalter, V., Ross, J. E., Catalano, L., & Fisch, S. M. (1990). Exploring language acquisition in children with a miniature artificial language: Effects of item and pattern frequency, arbitrary subclasses, and correction. *Journal of Memory and Langage, 29*, 591-610.

Brooks, P. J., Braine, M. D. S., Catalano, L., & Brody, R. (1993). Acquisition of gender-like noun classes in an artificial language: The contribution of phonological markers to learning. *Journal of Memory and Language, 32*, 76-95.

Bruhn, J., & White, L. (2000). L2 acquisition of Spanish DPs: the status of grammatical features. In S. C. Howell, S. A. Fish, & T. Keith-Lucas (Eds.), *Proceedings of the 24th Annual Boston University Conference on Language Development.* (Vol. 1, pp. 164-175). Somerville, Mass.: Cascadilla Press.

Carroll, S. (1989). Second language acquisition and the computational paradigm. *Language Learning, 39*, 535-594.

Carroll, S. E. (1999). Input and SLA: Adults' sensitivity to different sorts of cues to French gender. *Language Learning, 49*(1), 37-92.

Carroll, S. E. (2001). *Input and Evidence: The raw material of second language acquisition*. Amsterdam: John Benjamins.

Caselli, M. C., Leonard, L. B., Volterra, V., & Campagnoli, M. G. (1993). Toward mastery of Italian morphology: a cross-sectional study. *Journal of Child Language, 20*, 377-393.

Cleeremans, A., & Jiménez, L. (2002). Implicit learning and consciousness: A graded, dynamic perspective. In R. M. French & A. Cleeremans (Eds.), *Implicit Learning and Consciousness* (pp. 1-40). Hove: Psychology Press.

Corbett, G. (1991). *Gender*. Cambridge: Cambridge University Press.

Ellis, N. C. (1998). Emergentism, connectionism and language learning. *Language Learning, 48*, 631-664.

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

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

Frigo, L., & McDonald, J. L. (1998). Properties of phonological markers that affect the acquisition of gender-like subclasses. *Journal of Memory and Language, 39*, 218-245.

Gómez, R. L., & Gerken, L. (2000). Infant artificial language learning and language acquisition. *Trends in Cognitive Sciences, 4*, 178-186.

Hawkins, R. (2001). *Second Language Syntax: A Generative Introduction*. Oxford: Blackwell.

Holmes, V. M., & de la Batie, B. D. (1999). Assignment of grammatical gender by native speakers and foreign language learners. *Applied Psycholinguistics, 20*, 479-506.

Johnstone, T., & Shanks, D. R. (1999). Two mechanisms in implicit artificial grammar learning? Comment on Meulemans and Van der Linden (1997). *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 524-531.

Kelly, M. H. (1992). Using sound to solve syntactic problems: the role of phonology in grammatical category assignments. *Psychological Review, 99*, 349-364.

Knowlton, B. J., & Squire, L. R. (1996). Artificial grammar learning depends on implicit acquisition of both abstract and exemplar-specific information. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 169-181.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioural and Brain Sciences, 22*, 1-75.

Maratsos, M. P., & Chalkley, M. A. (1980). The internal language of children's syntax: The ontogenesis and representation of syntactic categories. In K. Nelson (Ed.), *Children's Language* (Vol. 2, pp. 127-214). New York: Gardner Press.

Marcus, G. F. (1999). Language acquisition in the absence of explicit negative evidence: Can simple recurrent networks obviate the need for domain-specific learning devices? *Cognition, 73*, 293-296.

Marcus, G. F., Vijayan, S., Bandi Rao, S., & Vishton, P. M. (1999). Rule learning in 7-month-old infants. *Science, 283*, 77-80.

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

Meulemans, T., & Van der Linden, M. (1997). Associative chunk strength in artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*, 1007-1028.

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

Pinker, S. (1984). *Language Learnability and Language Development*. Cambridge, Mass.: Harvard University Press.

Plunkett, K., & Elman, J. L. (1997). *Exercises in Rethinking Innateness: A Handbook for Connectionist Simulations*. Cambridge, Massachusetts: MIT Press.

Redington, M., & Chater, N. (1998). Connectionist and statistical approaches to language acquisition: A distributional perspective. *Language and Cognitive Processes, 13*, 129-191.

Saffran, J. R. (2001). The use of predictive dependencies in language learning. *Journal of Memory and Language, 44*, 493-515.

Shanks, D. R. (1995). *The Psychology of Associative Learning*. Cambridge: Cambridge University Press.

Skehan, P. (1998). *A Cognitive Approach to Language Learning*. Oxford: Oxford University Press.

Sokolik, M. E., & Smith, M. E. (1992). Assignment of gender to French nouns in primary and secondary language: A connectionist model. *Second Language Research, 8*, 39-58.

Sun, R., Merrill, E., & Peterson, T. (2001). From implicit skills to explicit knowledge: a bottom-up model of skill learning. *Cognitive Science, 25*, 203-244.

Taraban, R., & Kempe, V. (1999). Gender processing in native and nonnative Russian speakers. *Applied Psycholinguistics, 20*, 119-148.

Tomasello, M. (2000). The item-based nature of children's early syntactic development. *Trends in Cognitive Sciences, 4*, 156-163.

Towell, R., & Hawkins, R. (1994). *Approaches to Second Language Acquisition*. Clevedon: Multilingual Matters.

Ullman, M. T. (2001). The neural basis of lexicon and grammar in first and second language: the declarative/procedural model. *Bilingualism: Language and Cognition, 4*, 105-122.

Vigliocco, G., Antonini, T., & Garrett, M. F. (1997). Grammatical gender is on the tip of Italian tongues. *Psychological Science, 8*(4), 314-317.

Williams, J. N. (in preparation). Implicit learning of form-meaning connections .

Williams, J. N., & Lovatt, P. (2003). Phonological memory and rule learning. *Language Learning, 53*, 67-121.

Notes

[1] This language was derived from Italian. The determiners were derived from the Italian *il, i, un, dei, la, le, una*, and *delle* by systematically substituting consonants (l g, d t, n l). The nouns correspond to Italian nouns which end in *-e* in the singular and *-i* in the plural regardless of gender, e.g. *cliente* (masculine), *stazione* (feminine). Note that none of the participants in Experiments 1 and 2 (reported below) had any knowledge of Italian, and only two participants in Experiment 3 knew Italian at an intermediate level or better as an L2.

[2] The only difference was that in the experiment they also had to produce the inflection, whereas in the simulation the inflection was provided on the input. However, in the experiment the participants learned the correct plural inflections in the preliminary vocabulary learning phase, and not in the training phase of the main experiment. In any case the inflection provides no clue as to the correct determiner over and above the presence or absence of the plurality of the noun.

[3] A Root Mean Square error of 0.1 means that over all of the input patterns presented on a particular cycle the average difference between the actual output and the required output on each node was 0.1 units of activation. The point at which the correct output node was simply the most active occurred well before an RMS error of 0.1 was achieved.

[4] The Luce ^{ratio} was also used as a measure of network performance - the activation level of the correct output node divided by the sum of the activation over all output nodes. Perfect output would be indicated by a Luce ratio of 1.0. In this simulation the mean Luce ratio over 20 runs was 0.87.

[5] The number of hidden units was set to about two thirds of the number of input/output units so as to force the inputs through a reduced representational space, exerting pressure on the network to extract generalisations. Other simulations were performed with either 10 or 40 hidden units but the generalisation performance was similar to that reported here.

6 This experiment was run in collaboration with Helen East.