Rules and associations: hidden Markov models and neural networks in the psychology of learning

Visser, I.

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
6 Reaction times and predictions in implicit sequence learning

Testing predictions by the simple recurrent network

Abstract

The simple recurrent network model was used successfully by Cleeremans and McClelland (1991) to describe implicit sequence learning. In the simple recurrent network, reaction times are assumed to be inversely proportional to the activation value of the corresponding node in the network. This activation can be interpreted as the level of anticipation of the position of the next stimulus. Consequently, in a prediction task, the prediction can also be derived directly from the activities of the output nodes of the simple recurrent network. We investigate ability to predict subsequent stimuli and reaction times in an implicit sequence learning experiment. In addition to measuring reaction times, we assess subjects' ability to predict the position of subsequent stimuli. The simple recurrent network model does not predict a dissociation between prediction and reaction times. Our results are in accordance with this.

6.1 Sequence learning

Although implicit learning has been studied for over thirty years starting with (Reber, 1967), detailed modeling of implicit learning behavior has only recently been undertaken. This modeling has been based mainly on the simple recurrent network (SRN) (Elman, 1990). Cleeremans and McClelland (1991) have used this network to model subjects reaction times (RT) in an implicit sequence learning experiment. Dienes et al. (1999) have used a variant of the SRN to model transfer of implicit knowledge.

A number of different paradigms have been developed for studying implicit learning. One distinguishing characteristic of these paradigms is the way in which they establish the presence of implicit and/or explicit knowledge (see Jimenez et al., 1996, for discussion). The aim of this chapter is to study and compare two measures of implicit learning, namely reaction times (RT) and subjects' predictions of upcoming stimuli. In implicit learning research the sequential implicit learning paradigm, and the attendant use of reaction time as the primary measure of performance, has become increasingly popular (see for example Nissen and Bullemer, 1987; Cleeremans and McClelland, 1991; Seger, 1997). In the present study we use an augmented sequence learning paradigm which allows for online comparisons of reaction times and predictions.
In sequence learning, sequences of stimuli are presented to subjects. Unknown to subjects the sequences contain regularities that make them predictable. Regularity is brought about by either using finite state grammars (Cleeremans and McClelland, 1991; Jimenez et al., 1996; Jimenez and Mendez, 2001) or short repeating sequences (Nissen and Bullemer, 1987; Lewicki et al., 1987; Perruchet and Amorim, 1992; Frensch et al., 1994; Seger, 1997; Shanks and Johnstone, 1999). Responses are simply key presses on the keyword with a congruent mapping between stimulus and response. The effect of learning is established by comparing RTs on predictable trials, those trials that are generated say by a finite state grammar, with RTs on non-predictable trials, that is, purely random trials. Typically, in exit interviews subjects are unable to report the rules that were used to generate the sequence of stimuli. This is taken to indicate that subjects are not aware of what they have learned.

The outline of the chapter is as follows. In the next section we will outline the simple recurrent network model and derive predictions from it about RTs and prediction ability of subjects. In section 6.3 several measures of implicit/explicit knowledge are discussed as they are used in sequence learning experiments. In the next section our experiment and results is presented and in the final section the results are discussed.

### 6.2 The simple recurrent network and sequence learning

In the field of implicit learning, the work of Cleeremans and McClelland (1991) on the SRN model is highly relevant. The architecture of the SRN that they used is due to Elman (1990). The network, depicted in Figure 6.1, consists of three layers: input, hidden and output nodes. In addition, the network has a context or recurrent layer.

The SRN is used as a model of implicit learning in the following way. Stimuli are presented to the network at the input layer. Activation is propagated through the network resulting in activations at the output nodes. The context units keep a copy of the hidden unit activity that resulted from the presentation of the previous stimulus. Hence the context units provide the network with a memory trace of the previous stimuli. The SRN is trained using the backpropagation algorithm (Rumelhart and McClelland, 1986).

Cleeremans and McClelland (1991) use the SRN to model RT performance of subjects by assuming “that there is a linear reduction in RT proportional to the relative strength of the unit corresponding to the correct response” (p. 244). As we used a grammar with four letters (A, B, C and D) to generate our stimulus material, the SRN in Figure 6.1 has four input and four output units labeled A, B, C and D. The grammar and stimulus material are discussed in detail in the method section. The final step in using the SRN as a model of implicit learning is the interpretation of the activities of the output units. The activities of the

---

1Note that by assuming this relationship between RT and the node corresponding to the correct unit of the network, it is not possible for the network to predict an incorrect response. In practice this is not a problem since incorrect responses are very rare due to the simplicity of the task.
The simple recurrent network and sequence learning

Figure 6.1: Simple recurrent network.

Output units are normalized and then interpreted as relative probabilities of the next stimulus. The normalized activity of the correct output unit is thus used by Cleereman and McClelland (1991) as a measure of anticipation of the next stimulus position. Hence, a large activity of the output node corresponding to the next stimulus results in a fast response. This anticipation can also be used to make predictions of the upcoming stimulus location. The location corresponding to the output unit with the highest activity has the highest probability of being predicted. This means that the SRN model predicts a negative relationship between prediction performance and RTs, with RTs decreasing as prediction improves.

Since Cleereman and McClelland (1991) use the SRN to model implicit learning, it may seem strange to use the same model to model predictions. The prediction task is generally taken to be a measure of explicit knowledge. Prediction and generation tasks have been used in this fashion since the introduction of the generation task by Nissen and Bullemeur (1987) and others after them (Cleereman and McClelland, 1991; Jimenez et al., 1996; Perruche and Amorim, 1992; Shanks and Johnstone, 1999). In our interpretation, the activation values of the output nodes of the network are anticipation or expectation about upcoming events. Subjects are not necessarily conscious or aware of anticipations and expectations of upcoming events. In the behaviorist literature, it is shown that rats have expectations about upcoming events without our having to suppose that they are aware of these expectations (Rescorla and Wagner, 1972). In priming experiments similarly, people are not necessarily aware or conscious of their being primed for upcoming stimuli.

From the SRN we can derive predictions about performance on a sequence learning RT task which is alternated with prediction trials. The model predicts an inverse relationship between RTs and correct predictions such that fast RTs correspond to a high probability for making a correct prediction. Predictions and the speed of reacting are both taken as measures of subjects' expectations. The point here is that, according to the SRN, we do not expect a dissociation between predictions and RTs, whereas these measures are generally taken to represent different kinds of knowledge and hence researchers have tried to find such dissociations.
or demonstrate their absence (Perruchet and Amorim, 1992; Shanks and Johnstone, 1999). Of course, this is not to say that subjects do not acquire explicit knowledge in sequence learning experiments. In the next section we discuss in more detail different measures of implicit/explicit knowledge and dissociations and associations that are expected between these measures.

6.3 Prediction, generation and reaction time

Using the sequential implicit learning paradigm, Cleeremans and McClelland (1991) had their subjects respond to a sequence of stimuli generated by a probabilistic finite state grammar, which we describe in some detail below. To determine the effects of implicit learning, they measured reaction times, and found that these decreased as subjects got more training. Similar studies have been done where a generation task is used to assess explicit knowledge: subjects are required to predict the next stimulus at each trial instead of reproducing the stimulus as in the usual RT task (Nissen and Bullemer, 1987). Still other researchers have used a free generation task to assess how much knowledge subjects have gained from the training phase (Perruchet and Amorim, 1992; Shanks and Johnstone, 1999). In free generation subjects are required to produce a series of responses at will without feedback.

The reason for using different measures is to gain insight into the nature of subjects’ knowledge in sequence learning experiments. The direct measures, prediction and generation, but also verbal reports, are then taken as measures of explicit knowledge and are contrasted with the indirect measure RT performance. A dissociation between such measures is supposed to indicate that in fact different knowledge bases are constructed during sequence learning.

Certainly such dissociations between verbal reports and RT measures have been found in sequence learning and other implicit learning experiments. Usually subjects can not verbalize any knowledge of the sequence of stimuli that was presented to them (Cleeremans and McClelland, 1991; Reber, 1967, 1976). Dissociations, in sequence learning, between generation, prediction and recognition tasks on the one hand, and RTs on the other are disputed. It has been argued that verbal reporting is not very sensitive in assessing subjects’ explicit knowledge because subjects are reluctant to report knowledge of which they are not very confident (Perruchet and Amorim, 1992; Jimenez et al., 1996; Shanks and Johnstone, 1999). Therefore more sensitive tests have been proposed such as generation and recognition tests.

Perruchet and Amorim (1992) argue that the generation task, as it is usually administered with feedback, likewise is not very sensitive to detect explicit knowledge. Giving feedback on every trial makes the procedure difficult to compare with the normal SRT situation where no feedback is given during a block of trials. Moreover, feedback may result in intentional learning during the task. Perruchet and Amorim (1992, p. 787) therefore introduce the free generation test where no feedback is given at all and subjects are instructed to generate a series of trials “that looked like the series they saw in the preceding phases”.

Jimenez et al. (1996) use a continuous generation task to test for explicit knowledge in a sequence learning paradigm. In continuous generation “the next stimulus
as prescribed by the sequential structure is presented regardless of participants' prediction responses" (Jimenez et al., 1996, p. 952). They use a probabilistic set of rules, i.e. a finite state grammar, to generate their stimuli. This results in much more complex stimulus material than in the typical sequence learning experiment, where a short repeating sequence is used (Shanks and Johnstone, 1999; Lewicki et al., 1987, 1988; Perruchet and Amorim, 1992; Seger, 1997).

Perruchet and Amorim (1992) did several experiments to establish the onset of the availability of explicit knowledge by repeating their experiment with various lengths of the training sequence. They used a free generation task to show that explicit knowledge was available after only six repetitions of a ten trial sequence. Their free generation test, and recognition test and the continuous generation test as used by Jimenez et al. (1996), however, share one disadvantage. They are administered at the end of the experiment only after sequence learning is finished. It is desirable to have an online test of explicit knowledge which can then be used to monitor the availability of explicit knowledge throughout the experiment. As far as we know no such test has been used before in the literature on sequence learning. Here we use such a test.

In the present study, we use online cued prediction: RT trials are interspersed with trials at which subjects have to predict the next stimulus. These trials are indicated by displaying question marks on the presentation screen. After such a prediction trial the RT trials continue without feedback about the correctness of the prediction. In online cued prediction the RT trials and the prediction trials are very similar. Jimenez et al. (1996) emphasize the importance of this similarity to ensure the sensitivity of the test. There is no direct feedback on the prediction trials. In the instruction we emphasize that at prediction trials subjects should just type whatever key first comes to mind without pausing to think what it should be. We thereby ensure that the prediction trials have a minimal impact on the routine that subjects acquire when typing the responses to the RT trials. This procedure allows for online comparisons between the indirect RT measure of learning and the direct prediction measure. This should also enable us to find dissociations between these measures, if this occurs, early on in the experiment.

6.4 Experiment

To assess the relation between RTs and prediction of stimuli we carried out a sequence learning experiment in which a series of RT trials was interspersed with prediction trials. On the prediction trials subjects had to guess the location of the next stimulus. Jimenez et al. (1996) propose a generation task where subjects, having been presented with a stimulus, are required to predict the next stimulus location at each trial. In contrast, we used a procedure where subjects have to predict just one item at a time, after which the sequence learning RT trials are resumed. In so doing we ensure that the prediction trials have a minimal impact on the routine that subjects acquire in reproducing the sequence. Another important difference between our procedure and other generation tasks (e.g. Nissen and Bullemer, 1987), is that no feedback is given concerning the correctness of the
prediction. Rather, after subjects have made their prediction, the next stimulus of the sequence is presented with the same response-stimulus interval as between consecutive RT trials.

In the present experiment random sequences of trials are used to control for possible effects of motor training and to establish the effects of subjects’ acquisition of (implicit) knowledge of the grammar. We use a $2 \times 10$ within-subjects design. Two levels of grammaticality (grammatical and random) and ten levels of training practice. The prediction of an inversely proportional relation between RTs and prediction performance, as derived from the SRN, translates into three specific hypotheses. The first concerns the standard implicit learning effect, which should result in an interaction effect of condition and level of practice on RTs: if (implicit) learning occurs, RTs should decrease more on grammatical trials than on random trials. The second hypothesis concerns the (implicit) learning effect for prediction performance, which should result in an interaction effect of condition and level of practice: over time, prediction should improve for the grammatical trials, but not for the random trials. The third and most important hypothesis concerns the relation between correct predictions and RTs: RTs should be faster on trials leading to correct predictions than on trials leading to incorrect predictions.

### 6.4.1 Method

Subjects were given a four-choice serial RT task, consisting of a total of 5280 trials divided in 22 blocks of 240 trials each. The blocks were split into two sessions that were presented on two consecutive days. Unknown to subjects the sequence of stimuli followed a pattern that was generated using the finite state grammar that is described below. Because of the rather complex structure of the sequences generated with such a grammar we used the rather large number of 5280 trials.

#### Subjects

Twenty-four subjects, undergraduates at the Department of Psychology of the University of Amsterdam, participated in this experiment. They received either course credits or a fixed financial reward for their participation. In addition, they could earn financial bonuses for fast and accurate responding.

#### Display for reaction time and prediction trials

There were two types of trials: RT trials and prediction trials. On the RT trials subjects were required merely to react to the current stimulus by pressing the appropriate key (see Figure 6.2(a)). At prediction trials subjects were required to predict the next stimulus by pressing the appropriate key (see Figure 6.2(b)).

The alphabet of the grammar we used to generate stimuli has four letters. The letters were translated into screen positions as shown in Figure 6.2(a). At each RT trial an ‘x’ appeared in one of the quadrants of the computer display and the subjects were required to press the corresponding key on the numerical keypad of the keyboard. The keys 1, 2, 4 and 5 on the numerical keypad were used as the spatial configuration of the response keys is congruent with spatial configuration of
(a) Computer display for the RT trials. Subjects have to press the key corresponding to the quadrant of the screen where the $x$ is shown. The letters in the top-left corner of the quadrants were not part of the actual display.

(b) Computer display for the prediction trials. All quadrants contain a question mark and subjects have to choose the quadrant in which they expect the next stimulus to appear and press the corresponding key. The letters in the top-left corner of the quadrants were not part of the actual display.

Figure 6.2: Displays for RT and prediction trials.

the stimulus positions on the display. Subjects were instructed to hold their index finger over the middle of the four keys and press the appropriate key only with their index finger.

Figure 6.2(b) shows the display presented to subjects at prediction trials. In all four screen locations question marks were shown to indicate that the trial was a prediction trial and subjects were supposed to choose one of the four positions. Subjects were instructed to press any of the four keys at the prediction trials. In the instruction they were told that on prediction trials they had to press the key that “seemed right to them” and that they “should not pause to think about what the next stimulus should be but rather press the key that first comes to mind”.

Procedure

At the start of the experiment subjects were told that accuracy and speed were equally important. The experiment started with two small blocks of (twenty) trials to familiarize the subjects with the task. We did not record the responses to these trials.

Each of the 22 experimental blocks consisted of four subblocks in the following order: 20 random RT trials, 100 grammatical RT trials, 100 grammatical prediction trials and 20 random prediction trials. In the RT subblocks the subjects were required only to reproduce the stimuli. In the prediction subblocks, RT trials were interspersed with prediction trials. The response-stimulus interval was 300 ms.

The random RT subblocks consist of 20 trials in which the only constraint on the order of the stimuli was that no two consecutive stimuli were the same. This is standard practice in the sequence learning paradigm because it prevents undesired speed-up of responses due to priming effects (Nissen and Bullemer, 1987; Cleeremans and McClelland, 1991; Perruchet and Amorim, 1992; Shanks and Johnstone, 1999).
The grammatical RT subblocks consisted of a series of one hundred grammatical trials. The grammatical prediction subblocks consisted of a grammatical sequence of one hundred trials interspersed with prediction trials. On average each grammatical prediction subblock include 25.5 prediction trials, that is on average one in four trials was a prediction trial. In the random prediction subblocks subjects were presented with 20 trials in random order (except for the fact that there were never two consecutive identical trials) interspersed with an average of 4.5 prediction trials. To allow for a trial-by-trial comparison between prediction and reaction times, in block 11 the series of trials in both the grammatical RT subblock and the grammatical prediction block was identical. In this way it is possible to compare directly the RT on a given trial with the prediction made on the same trial within one block. In the last block, the order of the random and grammatical RT and prediction subblocks is reversed, first grammatical and then random, to check whether the order of the subblocks influences the responses.

**Stimulus material**

The sequence of stimuli in the grammatical subblocks was generated from the finite state grammar shown in Figure 6.3. Sequences are produced by this grammar in the following manner. First start in state #1 and randomly choose (with equal probability) one of the arcs leaving that state while noting the letter corresponding to the followed arc. This process is repeated in the next state by choosing a random arc. Note that in some states there is only one arc leaving that state and hence that arc is chosen. The process ends when state #7 is reached and the process starts again in state #1 to create strings of unbounded length. For example, for the grammatical prediction subblocks we generated a series of 100 trials from the grammar and selected random positions in this series that became prediction trials.

**Exit interviews**

After the experiment was completed all subjects were interviewed to assess whether subjects had acquired any explicit knowledge of the grammatical sequence. The questions were arranged to be increasingly specific. First they were asked about their thoughts on the purpose of the experiment. Second they were asked if they had anything to comment on the stimuli. Third they were asked whether they noted anything unusual in the sequence of the stimuli. Finally, after being told that there was a certain pattern in the stimuli, subjects were asked whether they had noted any pattern and if so, they were asked if they could point out the pattern on the screen or the keys.

**6.4.2 Results**

The data of one of the subjects were discarded, because the subject made too many errors (over 50%) in three consecutive blocks due to misplacing the index finger over the numerical keypad. Comparison of the last two blocks revealed that the order of the subblocks, random before grammatical or vice versa, did not significantly influence reaction times.
6.4 EXPERIMENT

Figure 6.3: Finite state automaton used to generate strings for sequence learning experiments. A string is formed by starting in state #1 and then randomly choosing one of the arcs leaving that state while noting the letter corresponding to that arc. Continue stepping from state to state until the end state #7 is reached; from there the process starts over again from state #1.

Reaction time trials

The first hypothesis states that RTs decrease more for the grammatical trials than for the random trials. In order to test this hypothesis, RTs were averaged over subjects and over two consecutive blocks. Grammatical RTs decreased from 404.7 ms (sd = 35.8) at the beginning of the experiment to 342.6 ms (sd = 32.7) at the end; random RTs decreased from 414.2 ms (sd = 43.2) to 370.3 ms (sd = 35.9). Mean RTs are displayed in Figure 6.4(a); bars around the means indicate standard errors. A $2 \times 10$ repeated measures ANOVA with two within factors, blocks (10 levels of practice) $\times$ grammaticality (2 levels), indicated a significant interaction between grammaticality and level of practice: as predicted, grammatical trial RTs decreased more over time than did random trial RTs, $F(9,198) = 3.87; p < 0.001$. The analysis also yielded significant main effects for grammaticality and block number: grammatical trial RTs were significantly smaller than the random trial RTs, $F(1,22) = 59.49; p < 0.001$, and RTs became faster over blocks for both grammatical and random trials, $F(3.78,198) = 25.75; p < 0.001$ (with Greenhouse-Geisser correction for deviance from sphericity (Stevens, 1996)).

Prediction trials

The correctness of prediction trials was scored as follows: a trial was scored as correct if the prediction made by the subject was identical to the actual successor in the series of stimuli that we generated. Due to the probabilistic nature of the finite state automaton used to generate the sequence of stimuli this means that in some cases the prediction is scored as incorrect although it is in fact a legal successor to the preceding stimuli according to the grammar. This is necessary to be able to compare percentages with the random blocks. The percentage of correct predictions
in grammatical subblocks increased from 36.0 % at the beginning to 52.2 % at the end of the experiment. The corresponding percentages for the random predictions are 30 and 34 %, respectively. The proportions of correct responses on predictions for both random and grammatical subblocks are displayed in Figure 6.4(b); bars around the means indicate standard errors. Note that the baseline for correct predictions is 33 %.

The second hypothesis states that prediction performance improves over time on the grammatical trials, but not on the random trials. A significant interaction between blocks (time) and grammaticality was found, \( F(7.9,198) = 2.25; p = 0.027 \), showing that the grammatical predictions improved more over time than did the random predictions. More specifically, there was no change over time for the random trial predictions when analyzed separately, \( F(1,228) = 0.845; p = 0.359 \), as was to be expected.

**Prediction and reaction time trials**

To compare performance on prediction and RT trials directly, in block 11 subjects were presented with an identical sequence of trials in the grammatical RT and prediction subblocks. The mean RT on correctly predicted trials was 360.96 ms (sd=49.6) and the mean RT on incorrectly predicted trials was 389.97 ms (sd=30.3). An ANOVA with one within factor (correct vs. incorrect) confirms that correct predictions are associated with faster RTs, \( F(1,22) = 6.44; p = 0.019 \).

**Trial by trial analyses**

To get a clearer picture of subjects' knowledge we carried out further analyses on both the RT and prediction data. We wanted to test whether subjects acquire a
better anticipation of the upcoming stimulus due to learning. To do so, subjects would have to encode the current state of the finite automaton (Jimenez et al., 1996). However, information about paths leading to a certain state suffices and hence we use this for further analyses. We focus on paths of length three since most of those paths determine the state of the grammar unambiguously.

First we computed the frequencies of all paths of length one to three from the series of stimuli presented to subjects. This includes both the grammatical and random subblocks. Second, for all paths of length three, from the data we computed the mean RT on the last stimulus of each path. For example, from the RT data we compute the mean RT for an A after AB, a C after AB and a D after AB. In fact only a C after AB is grammatical but from the random subblocks the other combinations are available as well. In the random subblocks a C after AB can also occur, albeit less frequently than in the grammatical subblocks, and so these RTs are also used in the analysis. Third, from the data we computed the conditional proportion of predicting each stimulus given the preceding two stimuli. We did this for both grammatical and non-grammatical continuations. For example, we counted the number of times subjects predicted an A after AB, a C after AB and a D after AB and we then normalized these for each context, AB in this case. Both the prediction and RT data were pooled over 4 consecutive experimental blocks. Restricting this computation to single experimental blocks would result in many empty cells because not all paths of length three necessarily occur in each prediction subblock. Finally we carried out multiple regression (stepwise backward) analyses to determine which frequencies, that is of single symbols, pairs or triples, were the best predictors of the differences in both the RTs and predictions. In Figure 6.5 the results of these analyses are plotted for the RT data. A session refers to aggregate data of four experimental blocks.

As can be seen in Figure 6.5 the proportion explained variance clearly increases with practice. Note that only the single symbol and pair frequencies entered into the analyses for sessions 1 to 4, indicating that subjects were basing their responses

Figure 6.5: Proportion explained variance of RTs by frequencies.
only on a single previous stimulus. Also plotted in Figure 6.5 are the coefficients that in the regression analyses of the single and pair frequencies. At sessions one through four the partial correlation from the triple frequencies with the RTs are plotted. At session five the coefficient from the regression analysis is plotted. The single frequencies become worse predictors over sessions whereas the pair and triple frequencies become better predictors as learning continues. An exception to this is session five, where the coefficient of the pair frequencies is smaller than in the first four sessions. This is due to the fact that in session five the triple frequencies are included in the analysis and the correlation between triple and pair frequencies is high (0.621, \( p < 0.0001 \)). At session five also the triple frequencies are included in the analysis indicating that at that point subjects based their responses on two preceding stimuli.

In Figure 6.6 a similar plot is shown for the proportion explained variance of the prediction data. Only the single symbol and pair frequencies are included in the analyses at all points; that is, the triple frequencies were at no point a significant predictor of subjects' predictions in the multiple regression analyses. This may be partly due to the fact that the correlation between pair frequencies and triple frequencies is high. Note that there is no clear pattern in the total proportion explained variance as was the case with the RTs. To get a clearer picture of the individual influences of the single symbol and pair frequencies, we also plotted the regression coefficients of single symbol and pair frequencies in Figure 6.6. As can be seen, the coefficient of the single symbol frequencies decreases whereas the coefficient of the pair frequencies increases. This indicates that, as learning continues, subjects base their predictions on pairs rather than on single symbol frequencies. Incidentally, this explains why prediction ability increases during training. For the triple frequencies separate simple regression analyses were done and the regression coefficients are also shown in Figure 6.6. All five regressions for triple frequencies were significant and the coefficients are clearly increasing in magnitude.
To investigate the role of anticipation in sequence learning we performed another regression analysis on the prediction and RT data together. In Figure 6.7 the pathdiagram of the model is shown. If subjects learn on the basis of frequencies, those frequencies determine their expectations about upcoming stimuli. Those expectations in turn determine whether subjects are able to respond fast in case of an RT trial.

Using LISREL (Jöreskog and Sörbom, 1999) we fitted five models on the aggregate data of four experimental blocks each as with the regression analyses above. We used the correlation matrix of the frequencies, the prediction data and the RT data, that is, a $5 \times 5$ correlation matrix in each model fit. In the model, we allowed measurement error for the predictions, $\varepsilon_p$ in Figure 6.7, since the prediction proportions are all based on small numbers of actual predictions made, making those data quite unreliable.

In Table 6.1 the estimated values are presented of the model parameters that are relevant to our hypotheses. First in Table 6.1 the four (standardized) regression coefficients are reported; $F_2$-$Pr$, for example, denotes the regression coefficient with the pair frequencies as predictor and the predicted proportion of a stimulus as dependent variable. The Pr-RT parameter is the regression coefficient from Pr to RT. In the next three columns of Table 6.1 the $\chi^2$ values are given along with the degrees of freedom and the associated $p$-values. In the final column the squared
Table 6.1: Parameter estimates for regression models of prediction and RT data combined.

<table>
<thead>
<tr>
<th>session</th>
<th>$F_3$-Pr</th>
<th>$F_2$-Pr</th>
<th>$F_1$-Pr</th>
<th>Pr-RT</th>
<th>$\chi^2$</th>
<th>df</th>
<th>p</th>
<th>$R^2$ for RT</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.00</td>
<td>0.34</td>
<td>0.82</td>
<td>-0.80</td>
<td>1.75</td>
<td>4</td>
<td>.78</td>
<td>.64</td>
</tr>
<tr>
<td>2</td>
<td>0.00</td>
<td>0.41</td>
<td>0.70</td>
<td>-0.91</td>
<td>2.39</td>
<td>3</td>
<td>.49</td>
<td>.84</td>
</tr>
<tr>
<td>3</td>
<td>0.00</td>
<td>0.49</td>
<td>0.66</td>
<td>-0.92</td>
<td>1.39</td>
<td>3</td>
<td>.71</td>
<td>.85</td>
</tr>
<tr>
<td>4a</td>
<td>0.00</td>
<td>0.58</td>
<td>0.56</td>
<td>-0.95</td>
<td>5.75</td>
<td>3</td>
<td>.12</td>
<td>.91</td>
</tr>
<tr>
<td>4b</td>
<td>0.18</td>
<td>0.47</td>
<td>0.56</td>
<td>-0.95</td>
<td>1.96</td>
<td>2</td>
<td>.38</td>
<td>.90</td>
</tr>
<tr>
<td>5</td>
<td>0.26</td>
<td>0.41</td>
<td>0.57</td>
<td>-0.94</td>
<td>0.74</td>
<td>2</td>
<td>.69</td>
<td>.87</td>
</tr>
</tbody>
</table>

multiple correlation of RT is given, which is similar to the $R^2$ in regression analysis.

All models fit the data well. In the models for sessions one to three, the regression coefficients from triple frequencies to Pr turned out to be not significant and hence were set to zero. For session four we included two models in Table 6.1. In the second model, referred to as 4b in Table 6.1, all the regression weights are included. Setting the regression weight $F_3$-Pr to zero in model 4b results in model 4a. A $\chi^2$ difference test shows that this regression weight is marginally significant, $\chi^2 = 3.79$ with $df = 1$, $p = 0.052$. That is, the fit of model 4b with the regression weight is marginally better than the fit of model 4a which does not have the regression weight.

As can be seen in Table 6.1 this pattern of results corroborates the other regression analyses. The influence of the single symbol frequencies remains significant throughout the five sessions, but diminishes while the influence of the pair frequencies and finally the triple frequencies in session four and five increases. For all five sessions the predictions turn out to be very good predictors for the RTs with regressions over 0.90 except for the first session.

Exit interviews

First subjects were asked about their thoughts on the purpose of the experiment. Six of the 24 subjects mentioned things like ‘trying to fathom the system’ and ‘learning about the system’. When asked whether subjects noticed anything particular in the sequence of stimuli, some of them felt there some ‘regularity’ in the sequence, but none of them could specify this. On this and the subsequent question whether they had noted any pattern in the stimuli, three subjects mentioned a combination of two screen positions that according to them occurred rather frequently. In all three cases this turned out to be the positions that, in their condition, corresponded with $AB$, that is, the loop between the two top right nodes in Figure 6.3. This combination indeed occurs frequently but is by no means the most frequently occurring pair of stimuli. For example, $DB$ occurs almost twice as frequently. In view of this it is hard to imagine that this pair of stimuli is solely responsible for the speed-up in RTs and the increasing probability of making correct predictions.
6.5 Discussion

In sequence learning both RTs and predictions have been used as a measure of performance. The results of this experiment show that, when measured simultaneously, it is possible to relate directly improvement in prediction performance to improvement in RT performance. The comparison shows that, as expected, fast RTs are indicative of the subjects’ level of anticipation of the next trial. The level of anticipation also results in differences in prediction performance. This study also shows that prediction is possible in a fairly complex rule system that can not be verbalized by subjects.

The results show that learning occurs: subjects give faster responses on grammatical trials than on random trials, and this effect becomes larger as learning continues. Secondly, subjects gradually get better at predicting subsequent stimuli. Thirdly, as expected, smaller RTs correspond with a better prediction of the subsequent stimulus. This can be seen from the significant correspondence between fast RTs and prediction ability in block 11. The close correspondence between prediction ability and RTs is even more pronounced in the regression models where it can be seen that the regression coefficients in the regression of RT on Pr are all very high. Only at the start of learning, in the first four blocks, is the regression coefficient below 0.90.

These results are in line with our expectations as derived from the SRN model, which does not predict a dissociation between these measures. The SRN successfully describes subjects’ growing sensitivity to dependencies between successive stimuli. Subjects first grow sensitive to first order frequencies, then to pair frequencies, then to triple frequencies. The regression analyses confirm this. The success of the SRN model is due to its ability to capture the ‘statistical constraints’ inherent in the sequence of stimuli (Cleeremans and Jimenez, 1998). The SRN model is used explicitly as a model of implicit learning and as such has been used successful (Cleeremans and McClelland, 1991; Cleeremans and Jimenez, 1998). As we have shown the SRN can also be used to explain prediction ability of subjects in the same experimental setting as is frequently used in sequence learning. It would therefore be strange to interpret the SRN as a model of explicit knowledge in this case.

In our interpretation, anticipation of the next stimulus determines both prediction ability and RTs. Since prediction ability is a direct measure of anticipation we used prediction ability as such in the regression model. Anticipation in turn is used as a predictor of RTs. All the models fitted the data very well. These findings support the interpretation that implicit learning is based on a growing sensitivity to the previous stimuli and their predictive value for the upcoming stimulus.

This still leaves room for the interpretation that both the increasing prediction ability and the faster RTs result from subjects’ gaining explicit knowledge instead of implicit knowledge. However, the strong association between these measures and the equally strong dissociation between these measures and subjects’ ability to verbalize their knowledge suggest that implicit learning has taken place. At least it points to the fact that in sequence learning, a single knowledge base is formed (see also Perruchet and Amorim, 1992; Shanks and Johnstone, 1999).