Psychologica Belgica
1997, 37-1/2, 89-113.

APPLICATION OF IMPLICIT KNOWLEDGE: DETERMINISTIC OR PROBABILISTIC?

Zoltán DIENES
University of Sussex, U.K.

Andreas KURZ Regina BERNHAUPT Josef PERNER
Universitaet Salzburg, Austria

This paper distinguishes two models specifying the application of implicit knowledge. According to one model, originally suggested by Reber (1967), subjects either apply sufficient knowledge to always produce a correct response or else they guess randomly (High Threshold Theory; subjects only apply knowledge when there is sufficient knowledge to exceed a threshold ensuring a correct response); according to the other model, suggested by Dienes (1992), subjects respond with a certain probability towards each item, where the probability is determined by the match between the items structure and the induced constraints about the structure (Probability Matching Theory; subjects match their probability of responding against their personal probability that the item belongs to a certain category). One parameter versions of both models were specified and then tested against the data generated from three artificial grammar learning experiments. Neither theory could account for all features of the data, and extensions of the theories are suggested.

Dienes and Berry (1997) argued that there is widespread agreement about the existence of an important learning mechanism, pervasive in its effects, and producing knowledge by unconscious associative processes, knowledge about which the subject does not directly have metaknowledge. Let us call the mechanism implicit learning, and the induced knowledge implicit knowledge. This paper will address the question of how implicit knowledge is applied.

Imagine an animal learning about which stimuli constitute food. Maybe its mother brings back different plants or animals for it to eat, and stops it from ingesting other plants or animals, which may be dangerous or poisonous. The details about what features go together to constitute something that is edible may be more or less complex; in any case, the animal eventually learns, let us say perfectly, to seek out only the edible substances in its environment. But before learning reaches this point, how should the animal behave towards stimuli about which it has only imperfect information and when the mother is not around to provide corrective feedback? One strategy would be to only

Correspondence concerning this article should be addressed to either Zoltan Dienes at Experimental Psychology, University of Sussex, Brighton, Sussex BN1 9QG, England. Electronic mail may be sent to dienes@epunix.susx.ac.uk; or Josef Perner at Institut fuer Psychologie, Universitaet Salzburg, Hellbrunnerstrasse 34, A-5020 Salzburg, Austria, e-mail josef.perner@sbg.ac.at.

ingest stimuli about which the animal's induced knowledge unambiguously indicates are edible; this would prevent unfortunate fatalities. On this scenario, implicit knowledge may have evolved in general to direct action towards a stimulus only when the implicit knowledge relevant to the stimulus unambiguously indicates the category to which the stimulus belongs. Partial information may not even be made available for controlling behaviour in case the animal uses the knowledge to sample poisonous substances. Thus, if the animal were *forced* to respond to some stimuli about which it had imperfect information, the animal would be reduced to random responding. We will call this the High Threshold Theory (HTT) - metaphorically, the knowledge lies fallow until it exceeds a high threshold.

Now imagine an animal learning about which locations are likely to contain food. This scene differs from the last in that a wrong decision is not so likely to have catastrophic consequences. If the animal's knowledge merely suggests that a location contains food, it still may be worth sampling in preference to another. There are two different states of affairs to be distinguished in this case:

First, different locations may have objectively different probabilities of containing food. The animal may come to have perfect knowledge of the different probabilities. It may seem rational for the animal to *always* choose the location with the highest probability, but this is not so if the probability structure of the environment is subject to fluctuations. Then it is beneficial to occasionally sample even low probability locations because it allows an assessment of whether the probability remains low. Also, because other animals will in general be competing for the food, and there will be more competitors at higher probability locations, it therefore pays to forage sometimes at low probability locations. In fact, animals do forage in different locations according to the probability of finding food there (Stephens & Krebs, 1986) - that is, they probability match. Similarly, in predicting probabilistic events, people like other animals are also known to probability match rather than respond deterministically. Reber and Millward (1968) asked people to passively observe binary events at a rate of two per second for two or three minutes. When people were then asked to predict successive events they probability matched to a high degree of accuracy (to the second decimal place; see Reber, 1989, for review).

Second, different locations may have imperfectly known probabilities of containing food. This is the case in which we are interested; how does the animal apply its imperfect knowledge? The knowledge may be imperfect because the locations have not yet been sampled frequently, or because the location is new and the animal does not know how to classify it. Consider an animal faced with choosing one of two locations. In the absence of certain knowledge about the objective probability structure, one strategy would be to search each location according to the estimated probability. For example, in the

absence of any information, the animal would search each location with equal probability, and adjust the search probabilities as information was collected. Or if features of the locations suggested different probabilities, search could start at those estimated probabilities. The true probabilities could be honed in on more quickly if the animal starts near their actual values. We will call the hypothesis that the animal responds to stimuli with a probability that matches the stimuli's expected probabilities Probability Matching Theory (PMT). If we allow probabilities to apply to singular events, then animals may respond probabilistically to stimuli for which there are no objective probabilities other than 0 or 1. For example, in classifying an edible stimulus as either being palatable or unpalatable, the animal may ingest it with a probability determined by the likelihood that the stimulus is palatable. In this case, the probability of the animal ingesting a stimulus may be interpreted as the animal's 'personal probability' that the stimulus is edible.

We have outlined two hypotheses about the way in which implicit knowledge could be applied: HTT and PMT. It may be that implicit knowledge is applied in different ways according to context; for example, according to the cost of a wrong choice, as in the examples of ingesting stimuli and searching locations given above. Or it may be that evolution has for economy produced a mechanism with a single principle of application. We will investigate these two hypotheses - HTT and PMT - in one context, that of people implicitly learning artificial grammars. In that field, both hypotheses have been suggested as accounting for human performance.

In a typical artificial grammar learning experiment, subjects first memorize grammatical strings of letters generated by a finite-state grammar. Then, they are informed of the existence of the complex set of rules that constrains letter order (but not what they are), and are asked to classify grammatical and nongrammatical strings. In an initial study, Reber (1967) found that the more strings subjects had attempted to memorize, the easier it was to memorize novel grammatical strings, indicating that they had learned to utilize the structure of the grammar. Subjects could also classify novel strings significantly above chance (69%, where chance is 50%). Reber (e.g., 1989) argued that the knowledge was implicit because subjects could not adequately describe how they classified strings (see Dienes & Perner, 1996, and Dienes & Berry, 1997, for further arguments that the knowledge should be seen as implicit).

Reber (1967, 1989) argued for a theory of implicit learning which combined HTT with the Gibsonian notion of veridical information pick-up. Specifically, Reber argued that implicit learning results in underlying representations that mirror the objective structures in the environment. In the case of artificial grammar learning, he argued that the status of a given test item would either be known (and would thus always be classified correctly); or the status was not known, and the subject guessed randomly. On this assumption, any knowledge

the subject applied was always perfectly veridical; incorrect responses could only be based on failing to apply knowledge rather than on applying incorrect knowledge. Reber used these assumptions by testing each subject twice on each string without feedback. If the probability of a given subject using perfect knowledge of the *i*th string is k_i then the expected proportion of strings classified correctly twice, once, or not at all by that subject is determined by the following equations:

the proportion of strings classified twice correctly = $p(CC) = k + (1-k)*0.5^2$

the proportion of strings classified correctly just once = $p(CE) + p(EC) = (1-k)* 0.5$

the proportion of strings never classified correctly = $p(EE) = (1-k)* 0.5^2$

where k is the average of the k_is.

Under this model, the values of p(CE), p(EC) and p(EE) averaged across subjects should be statistically identical and lower than p(CC). If p(EE) is greater than p(CE) or p(EC), then this is evidence that subjects have induced rules that are not accurate reflections of the grammar; the incorrect rules lead to consistent misclassifications. Reber (1989) reviewed 8 studies in which subjects in the learning phase were exposed to grammatical stimuli presented in a random order so that the grammatical rules would not be salient, and in which in the test phase subjects were tested twice. When subjects were asked to search for rules in the learning phase, p(EE) was on average .22 and the average of p(CE) and P(EC) was .13. That is, when subjects were asked to learn explicitly they induced substantial amounts of nonrepresentative rules. When subjects were asked to memorize strings in the learning phase, p(EE) was on average .16 and the average of p(CE) and p(EC) was .12. That is, when subjects were encouraged to learn implicitly, subjects only applied knowledge when it was largely (even if not entirely) representative, consistent with Reber's claim that knowledge only applied when it unambiguously indicated a correct classification.

In contrast to Reber, Dienes (1992; Dienes & Perner, 1996) suggested a connectionist model of artificial grammar learning that applied its knowledge according to PMT. The model became sensitive to the actual statistical structure of the environment, and to this degree embodied Reber's notion of representational realism. However, the network was a two-layer connectionist network and so could only become sensitive to first order statistical structure, and not higher orders; in this way, it's knowledge did not entirely mirror that of the environment. During the learning phase the network was trained to predict each letter from every other letter in the learning string. In the test phase, it attempted the same prediction. The activation of an output unit was determined by the correlation between the actual and predicted letters in the test string. The output activation directly determined the probability that the network would

give a 'grammatical' response. Thus, the more each letter in the test string could be accurately predicted from the other letters, the greater was the probability of the network saying 'grammatical'.

Dienes (1992) showed that for a set of stimuli that had been used in the past in human experiments by both Dulany, Carlson, and Dewey (1984) and Dienes, Broadbent, and Berry (1991), the proportion of subjects classifying each test string correctly was a reliable feature of the data across different experiments. The proportions for different strings varied between 37% and 82% with a good spread inbetween. Reber's model would interpret this variation in proportions as variation in the proportion of subjects who knew each string (with the values below 50% being due to the occasional nonrepresentative knowledge induced by subjects). Dienes interpreted the variation not as variation across subjects but as a variation across items in the probability of saying 'grammatical' (it was assumed that this probability was constant across subjects). In other words, subjects could be 'probability matching' (in quotes because there are no objective probabilities to match - each test string either is or is not grammatical). The model incorporated the assumption of 'probability matching' in that the activation of the output unit codes the degree that the test string satisfies the internal constraints induced by the grammar, and hence a sort of personal probability that the string is grammatical.

Dienes (1992) showed that the connectionist network could successfully predict the probability of each test string being called grammatical (i.e. the proportion of subjects who called it grammatical), the overall classification performance of subjects, and the p(CC), p(CE), and p(EE) values averaged over subjects and strings. That is, Dienes showed that a model incorporating the assumptions of PMT could account for the same data that Reber (1989) had argued supported his HTT. The closeness of p(CE) and p(EE) could only be produced in the PMT model by having some proportion of items with a probability correct, p, well below .50. If $p > .50$, then $p(CE) > p(EE)$ since $p(1-p) > (1-p)^2$.

In summary, the data published so far do not discriminate the two theories of application of implicit knowledge, PMT and HTT. The aim of this paper is to directly test the two theories by testing each subject three times on a set of test strings and looking at the distribution of correct responses for each string separately.

If people are tested three times on each string instead of twice, the Reber model makes the following predictions, where X is the proportion of people who *know* the string:

the proportion of people classifying an item three times correctly =

$$p(3Cs) = X + (1-X)*0.5^3$$

the proportion of people making two correct classifications and one error =

$$p(2Cs, 1E) = (1-X)*3*0.5^3,$$

the proportion of people making two errors and one correct classification =

$$p(1C, 2Es) = (1-X)*3*0.5^3$$

and the proportion of people making three errors =

$$p(3Es) = (1-X)* 0.5^3.$$

According to the PMT model, the corresponding equations are

$$p(3Cs) = p^3$$
$$p(2Cs, 1E) = 3*p^2*(1-p)$$
$$p(1C, 2Es) = 3*p*(1-p)^2$$
$$p(3Es) = (1-p)^3$$

where p is the probability of responding correctly to the string.

Both models contain just one parameter which can be estimated from the overall proportion, P, of correct classifications for each item. For the Reber model, X can be estimated from the relationship $X + (1-X)*0.5 = P$ (thus, $X = 2*P - 1$), and for the PMT model p is estimated directly by P.

In some cases, the models make qualitatively similar predictions. For example, the Reber model predicts that p(3Cs) should be greater than p(2Cs, 1E) given that $P > .70$. Similarly, the PMT model predicts that p(3Cs) should be greater than p(2Cs, 1E) given that $P > .75$. Further, both models predict that for Ps close to .50, the distribution of responses should approximate a binomial with p=0.50.

The models also make qualitatively different predictions. For the Reber model, regardless of P, p(2Cs,1E) should always be equal to p(1C,2Es). For the PMT model, these probabilities will differ except in the special case that $p = 0.5$. If $p > 0.5$ then p(2Cs,1E) will be greater than p(1C,2Es). Also for the Reber model, regardless of P, p(1C,2Es) should always be equal to 3*p(3Es). For the PMT model, p(1C,2Es) will be greater than 3*p(3Es) whenever $p > 0.5$.

The models can also be fitted quantitatively to the actual distributions obtained. These qualitative and quantitative predictions are explored in three experiments.

Experiment 1

Method

Participants. 127 first year psychology majors from the University of Salzburg participated in a class experiment. There were 33 men and 94 women with ages ranging from 18 to 37 years.

Materials. The finite state grammar and stimuli used were the same as used by Reber and Allen (1978), and later by Dienes (1992), amongst others.

Learning items. The learning items were 3 to 5 letters long, and composed of the letters M, R, X, V, and T. The items were projected overhead so that the height of letters subtended about 1/2° to 2° of visual angle (depending on the distance in the auditorium). In addition to visually displaying the items, the experimenter spelled them out aloud.

Test items. Reber and Allen's (1978) set of test items was used with the following modifications. Any items that were part of the learning set and any items of less than 4 letters were omitted. Of the remaining items, 6 grammatical and 6 ungrammatical items were selected as "repetition items". The selection criteria were to find in each category 3 items 5 letters long and 3 items 6 letters long for which the classification accuracy in Dienes (1992) was around 67%. The eventually selected items had accuracies that varied between 57% and 78%. The final test list consisted then of 63 items: 27 non-repeated strings (11 grammatical and 16 ungrammatical) and 12 strings that were repeated 3 times each (making another 36 items). These items are shown in Table 1. Each of the repeated sequences was placed once in each successive third of the whole set. This placement was random with the restriction that for each item there were at least 10 intervening items before its repetition.

Procedure

Learning. The class of 127 students was divided into three groups on the basis of seating arrangements. This resulted in 45 students in the No Exposure Group (no learning experience), 33 in the Single Exposure Group (single exposure to the learning set) and 49 students in the Double Exposure Group (two exposures to the learning set). After a general introduction that students are to participate in a memory experiment with three different groups, members of the No Exposure Group were sent out of the auditorium. The two groups remaining were given a recording sheet and received the following instructions:

> "I will now show you a list of letter strings which you should try to remember. To help you, I will show only 4 items at a time for a few seconds. Try to memorise these items and then reproduce them on your sheet of paper from memory. Do not start writing before the items have been covered up again! I will then show you the items again. Mark how many of them you had written down correctly without a single letter wrong. Then I'll show you the next four items, and so on."

The timing of items was determined by spelling out each item at a pace of 3 to 5 sec per item. The exposure of a group of 4 items lasted, therefore, about 16 seconds on average. After the end of the 20 learning items the members of the Single Exposure Group handed in their recording sheets and then left the auditorium with instructions not to talk about the learned items among themselves or to the members of the No Exposure Group.

Table 1
Stimuli Used in Experiment 1

repeated grammatical items
MTTTVT
MTVRXV
VXRRR
VXTTV
MVRXR
MTTVRX
repeated nongrammatical items
MMVRX
VVXRM
TVTTXV
XVRXRR
MTRVRX
VXRVM
nonrepeated grammatical items
VXTV
MTTTV
VXRM
VXRRRM
VXTTTV
MTVRXR
VXVRXR
MTVT
MVRXRM
MTTV
MVRXM
nonrepeated nongrammatical items
MXVRXM
TTVT
XRVXV
MXVT
MTVV
MVRTR
MTRV
VXRRT
VXMRXV
TXRRM
MTTVTR
RRRXV
MTVRTR
VRRRM
XTTTTV
VXRT

The remaining students (Double Exposure Group) were instructed to turn over their recording sheet for a second run through the learning list. After completion the recording sheets were collected, the other two groups were called back to the auditorium and the lists with test items distributed.

Test. The list of 63 test items was presented in one column stretching over the two sides of an A4 sheet. Students were given the following explanation and instructions:

> "The sequences you've just seen were produced by a system of rules, i.e., the grammar of an artificial language. I will now show you further sequences of this kind and you should decide for each item whether it belongs to the same artificial language (that is, has been produced by the same system of rules as the items you have studied) or not."

For the No Exposure Group the following addition was made:

> "For you these instructions may seem nonsensical, since you have not yet seen any such sequences. However, it is often possible to just use one's feeling to classify items one way or the other: Some sequences just feel more natural than others. Try to do your best."

All students were then told to put next to each item in column "A" their answer "y" if they think the item is grammatical or "n" if they think it is not. They should also provide an estimate of their confidence in column "C" by putting 50% if they are completely unsure of their judgement and 100% if they are absolutely certain or any percentage in between depending on the strength of their confidence.

At the very end students were asked whether (1) they had ever participated in a similar artificial grammar learning experiment before, whether (2) they thought they had formed any rules about the strings, and whether (3) they had noticed anything else about the strings (the hope was that this would bring students who had noticed the repetition of items to remark on it).

Results

Over all repeated and nonrepeated items, the No Exposure Group had a classification performance of 54% (SD=8%), the Single Exposure Group had a classification performance of 62% (SD=8%), and the Double Exposure Group had a classification performance of 66% (SD=8%). The groups differed significantly, $F(2, 124) = 24.46$. $p < .005$. That is, the amount of exposure in the learning phase influenced subsequent classification performance.

Appendix 1 shows the data for each repeated test string individually, displaying the number of subjects getting each string correct three times, twice, once, or not at all, and the fits of the PMT and HTT models. Some strings had Ps below .50, sometimes substantially so. This poses a problem for the HTT

model outlined in the introduction because it only allowed subjects to know a string correctly or else guess. Thus, it was assumed that in cases where P was below .50, there was a proportion X of subjects who "knew" a string incorrectly; i.e. a proportion X of subjects always gave the wrong response and the remaining (1-X) of subjects guessed randomly. For all other Ps, the assumptions given in the introduction were applied.

Qualitative comparisons. Qualitatively consistent with both theories all items from the learning groups with a P of .70 or greater had higher p(3Cs) than P(2Cs, 1E).

For $P > 0.5$, PMT predicts that $p(2Cs,1E) > p(1C,2Es)$; HTT predicts that the two should be identical. In fact, for the 15 cases from the learning groups that had $P > .60$, all had $p(2Cs,1E) > p(1C,2Es)$, competitively supporting PMT over HTT.

For $P > 0.5$, PMT predicts that $p(1C,2Es) > 3*p(3Es)$; HTT predicts that the two should be identical. In fact, for the 15 cases from the learning groups that had $P > .60$, only 4 had $p(1C,2Es) > 3*p(3Es)$, disconfirming both theories.

Quantitative comparisons. The fit between the models and the data was determined for each item in each condition by computing a chi-square based on the observed and predicted numbers of subjects classifying the item correctly exactly three times, twice, once, or not at all (see Appendix 1). Summing over items, the total chi-square values for the learning groups for PMT and HTT are 300.91 (df=72) and 135.96 (df=72), respectively, $p << .005$ in both cases. That is, the data departed significantly from both models, with a better fit for HTT rather than PMT. By a sign test over items, the fit was significantly better for HTT rather than PMT, $p = .0005$.

In the No Exposure Group, the total chi-squares for PMT and HTT were 494.69 (df=36) and 273.72 (df=36), respectively; in the Single Exposure Group, the total chi-squares for PMT and HTT were 188.55 (df=36) and 62.31 (df=36), respectively; and for the double exposure group they were 112.36 (df=36) and 73.65 (df=36) respectively. Notice that for PMT, chi-squares approximately halved with each successive exposure; that is, the fit of the model became markedly better as learning progressed. For PMT, the Single Exposure Group was significantly better than the No Exposure Group, $p = .006$ by sign test over items; but the Double Exposure Group was not significantly better than the Single Exposure Group, thus it is not clear that this increase in fit would apply to other items from the grammar. For HTT, the fit improved with one exposure of the learning items rather than no learning ($p = .006$ by sign test), but levelled off after one exposure.

What is the reason for the HTT model's superiority? In 23 of the 24 cases in the learning conditions, there were more consistent responders than PMT

predicted; i.e. both p(3Cs) and p(3Es) were higher than predicted. HTT fared much better, underpredicting p(3Cs) in only 13 out of 24 cases, and underpredicting p(3Es) in only 18 out of 24 cases.

In the No Exposure Group, both PMT and HTT underpredicted both p(3Cs) and p(3Es) in all 12 cases. In the No Exposure Group, the distribution consisted mostly of consistent responders, inconsistent with both models.

Experiment 2

The main findings of Experiment 1 are that PMT rather than HTT better predicted a greater proportion of subjects getting just two rather than just one trial correct, but that HTT provided the better overall quantitative fit by virtue of fitting p(3Cs) more closely than PMT. Both theories tended to underpredict p(3Es). The aim of experiment 2 was to replicate these findings when subjects were tested individually rather than as a group.

Method

Participants. Thirty students (three men and 27 women) from the University of Salzburg volunteered for this study. They were aged between nineteen and fifty-seven (M=27.47, SD=8.86).

Stimuli. The artificial grammar and the training and test strings were the same ones as used in Experiment 1 with some modifications:
Letters used were M, R, T, X, and Z. The original letter V was replaced by Z because V was hardly distinguishable from U on the display. The set of twenty training strings remained otherwise unchanged. Test strings consisting of three letters were not used so that there remained twenty-one grammatical and twenty-two nongrammatical strings. Therefore one grammatical string (MTZR) was added. The four grammatical strings in the original test set that appeared in the training set were replaced (MZRXZT was replaced by MZRXZR, ZXZT by ZXZR, ZXZRX by ZXTZR, and ZXZRXZ by ZXZRXT).

Twelve strings (six grammatical and six nongrammatical ones) were presented three times in the test phase (MTTTZT, MTZRXZ, ZXRRR, ZXTTZ, MZRXR, MTTZRX; MMZRX, ZZXRM, TZTTXZ, XZRXRR, MTRZRX, ZXRZM). In all, the test set consisted of sixty-eight strings.

A Power Basic program controlled the display of instructions, stimuli, and collection of responses. For each participant the training and the test strings were randomized anew by the computer so that no participant had the same sequence of training or test strings. The repeated test strings were positioned so that one instance appeared in each successive third of the test sequence. Within

each third, the item was positioned randomly with the restriction that at least two intervening items must separate repetitions of the same item.

Procedure.

All participants were individually tested. Participants were initially recruited to take part in a 15-min learning experiment. They were taken to an office and placed in front of a computer. Then all instructions were given on the screen. After having welcomed the participant and having asked demographic data (sex, age) the computer displayed:

> "This experiment is subdivided into two sections. In the first section you are to learn a number of items. Each item is made of several letters, e.g., 'ABCDE'. The items appear separately one after the other in the middle of the screen. Each item stays visible only for a short time before the next item follows. During this time you are to learn the item by spelling it out loudly. If you have any questions about this section please ask the experimenter now, if not then concentrate your attention on the middle of the screen and push the space bar to begin!"

After the participant had pushed the space bar the training strings were displayed as described. At this time the participants neither knew that the strings were produced by an artificial grammar nor what they had to do in the second part of the experiment. Each string appeared for three seconds on the screen in white letters on a black background. Then it disappeared and after a break of one second the next string appeared and so on.

Having completed the training phase the participants were told:

> "You have now learned a number of items. Each of these letter strings is based on the same 'grammar', i.e., the order of the letters within each item was not determined by chance but by some underlying structure. In the second section of this experiment you are to judge items that you have not yet seen. An item is 'grammatical' if it is based on the same structure as the ones you learned. Otherwise the item is 'not grammatical'. You have to decide whether an item is grammatical or ungrammatical and then you indicate on a percentage scale how confident you are in your decision. Choose 50% if you think it was a pure guess. If in your opinion your judgement goes beyond guessing choose a correspondingly higher percentage. Make the required indications with the 'mouse' by moving the arrowhead into the chosen area and then pushing the left mouse key."

The computer demonstrated graphically the described procedure. Then the participants were given the opportunity to practise handling the computer mouse by dealing with example strings consisting of random arrangements of the letters A, E, I, O, and U. They could decide themselves wether they wanted

to practise or not. Practising could be broken off after each example. As soon as the participant decided to have practised enough the computer displayed:

> "If you have any questions about this section please ask the experimenter now, if not take the 'mouse' into your right hand and push the space bar with your left hand to begin! (Dealing with all items takes about 10 min.)"

After the participant had pushed the space bar the first test string was displayed. The test string itself was placed in the middle of the screen (as the training strings before) in white letters on a black background. The answer choice "grammatical" was written above "not grammatical". The "mouse pointer" was placed between these two words. After the participant's response the two answers choices (but not the test string) vanished and a scale (in white colour) appeared under the test string. This scale consisted of six equally spaced frames around the percentages 50, 60, 70, 80, 90, and 100. These frames were connected with straight lines. The "mouse pointer" was placed below the middle of the scale. The "mouse-click" could be set within each frame or on the lines between frames. After the response the screen was cleared (black screen) and after one second the next test string appeared. The participants did not get feedback about their classifications. They had as much time as they wanted for their answers.

After the sixty-eight test strings were completed the participants were asked:

> "Please answer the following question:
> In your opinion: How many of the items did you judge correctly?
> (Hint: If you only guessed then it should be about half of all items.)
> Of the 68 items you judged correctly: ——? (any number between 34 - 68)"

The participants could choose a number only between thirty-four and sixty-eight. Other values were not accepted by the computer. After this last input the computer thanked the participant and said goodbye.

Results

Overall, the classification performance was 60% (SD=7%). Appendix 2 show the data for each repeated test string individually, displaying the number of subjects getting each string correct three times, twice, once, or not at all, and the fits of the PMT and HTT models.

Qualitative comparisons. The overall pattern was the same as experiment 1.

Qualitatively consistent with both theories, all items from the learning groups with a P of .70 or greater had higher p(3Cs) than P(2Cs, 1E).

For $P > 0.5$, PMT predicts that p(2Cs,1E) > p(1C,2Es); HTT predicts that the two should be identical. In fact, for the 7 cases that had $P > .60$, 6 had p(2Cs,1E)

> p(1C,2Es), competitively supporting PMT over HTT.

For P > 0.5, PMT predicts that p(1C,2Es) > 3*p(3Es); HTT predicts that the two should be identical. In fact, for the 7 cases that had P > .60, 6 had p(1C,2Es) < 3*p(3Es), inconsistent with both theories.

Quantitative comparisons. The total chi-square values for PMT and HTT were 138.18 (df=36) and 69.29 (df=36), respectively, $p < .005$ in both cases. That is, the data departed significantly from both models, but the fit was closer for HTT rather than PMT ($p = .039$ by sign test).

In terms of consistency of responding, in 11 out of 12 items, PMT underpredicted both p(3Cs) and p(3Es). HTT fared better with p(3Cs), underpredicting p(3Cs) in exactly 6 out of 12 cases, but underpredicting p(3Es) in 10 out of 12 cases.

Experiment 3

Experiment 1 had found model fits to improve as amount of exposure in the learning phase increased. Experiment 3 attempted to replicate Experiment 1 with more exposure in the training phase.

Method

The same method as Experiment 1 was followed with the following exceptions. Different groups of subjects were exposed to the training list once, twice, or three times (the Single Exposure, Double Exposure, and Triple Exposure Groups, respectively). Moreover, to intensify the learning experience, four new learning items (namely, VXTTTV, MTTTV, VXVRXR, MVRXM) were added at the end of the second repetition, and another four (VXRRRM, VXTV, MTVRXR, MVRXRM) at the end of the third repetition.

Results

Overall, the Single Exposure Group had a classification performance of 61% (SD=7%), the Double Exposure Group had a classification performance of 66% (SD=7%), and the Triple Group had a classification performance of 69% (SD=8%). The difference between the groups was significant, $F(2, 134) = 12.08$, $p < .01$. That is, the amount of learning exposure influenced subsequent classification performance.

Appendix 3 shows the data for each repeated test string individually,

displaying the number of subjects getting each string correct three times, twice, once, or not at all, and the fits of the PMT and HTT models.

Qualitative comparisons. Qualitatively consistent with both theories, all 17 items with a P of .70 or greater had higher p(3Cs) than p(2Cs, 1E).

For P > 0.5, PMT predicts that p(2Cs,1E) > p(1C,2Es); HTT predicts that the two should be identical. In fact, for the 23 cases that had P > .60, 21 had p(2Cs,1E) > p(1C,2Es), competitively supporting PMT over HTT.

For P > 0.5, PMT predicts that p(1C,2Es) > 3*p(3Es); HTT predicts that the two should be identical. In fact, for the 23 cases that had P > .60, 13 had p(1C,2Es) < 3*p(3Es), this time supporting HTT over PMT.

Quantitative comparisons. The total chi-square values for PMT and HTT are 402.48 (df=108) and 219.38 (df=108), respectively, $p < .005$ in both cases. That is, the data departed significantly from both models, with a better fit to HTT rather than PMT ($p = .039$ by sign test over items).

In the Single Exposure Group, the total chi-squares for PMT and HTT were 185.6 (df=36) and 99.09 (df=36), respectively, for the double exposure group they were 135.47 (df=36) and 67.98 (df=36) respectively, and for the triple exposure group they were 81.43 (df=36) and 52.31 (df=36) respectively. For PMT, the increase in fit from one to two exposures was not significant over items by sign test; nor was the increase from one to two exposures. For HTT, the changes were not significant either.

In quantitative comparisons, HTT did better than PMT because of its ability to predict consistent responders. In 30 of the 36 cases, PMT underpredicted both p(3Cs) and p(3Es). HTT on the other hand, underpredicted p(3Cs) in only 16 out of 36 cases, and underpredicted p(3Es) in 23 out of 36 cases.

Discussion

Experiment 3 replicated the finding that PMT rather than HTT better predicted a greater proportion of subjects getting just two rather than just one trial correct, but that HTT predicted p(3Cs) more closely than PMT. Experiment 3 further showed that the fit of both models improved with increasing learning (although not consistently for all items).

General Discussion

This paper has distinguished two models specifying the application of implicit knowledge. According to one model, originally suggested by Reber

(1967), subjects either apply sufficient knowledge to always produce a correct response or else they guess randomly (High Threshold Theory; subjects only apply knowledge when there is sufficient knowledge to exceed a threshold ensuring a correct response); according to the other model, suggested by Dienes (1992), subjects respond with a certain probability towards each item, where the probability is determined by the match between the items structure and the induced constraints of the grammar (Probability Matching Theory; subjects match their probability of responding against their personal probability that the item is grammatical). One parameter versions of both models were specified and then tested against the data generated from three artificial grammar learning experiments. The experiments tested subjects three times on a set of test items; each model made distinct predictions about the distributions of subjects classifying correctly exactly three times p(3Cs), two times p(2Cs, 1E), one time, p(1C, 2Es) or not at all p(3Es).

PMT, but not HTT, successfully predicted that p(2Cs, 1E) should be greater than p(1C, 2Es) given that $p > .50$. However, PMT could not predict the amount of consistency in the data; unlike HTT, it systematically underpredicted p(3Cs). Both theories tended to underpredict p(3Es).

What could be the reasons for both theories failing to predict how consistent subjects are in responding? First we will consider a reason in terms of why the experimental data may not have been an appropriate test of the models. The predictions derived from the models assume that the successive tests of each subject are independent trials. If subjects use implicit or explicit memory of previous responses to the same item, this may spuriously increase consistency. One way of testing whether subjects use memory is based on the assumption that the immediately preceding presentation of the same item should have more memory influence than the one before that, i.e., if subjects are using memory then $p(E_3/E_2C_1)$ should be greater than $p(E_3/C_2E_1)$, where E_j refers to making an error on the jth trial. These probabilities are expected to be identical on both PMT and HTT accounts. Using the data from Experiment 3, $p(E_3/E_2C_1) = .57$ (based on 159 observations) and $p(E_3/C_2E_1) = .39$ (124 observations), the difference is significant, $\chi^2 = 8.92$, $df = 1$, $p < .005$. That is, there is evidence that memory did influence subjects' responses. Future research could try to test subjects with repeated items separated by more nonrepeated items than used in the current experiments, to try to minimize subjects' memory of previous responses.

Independently of subjects' memory for responses, there are also natural extensions of HTT and PMT that would produce greater consistency. HTT could be plausibly extended to produce higher p(3Es). In particular, a proportion, X, of subjects might always classify a string correctly, and a proportion, Y, might have induced nonrepresentative knowledge and always classify the same string incorrectly. The remaining (1-X-Y) subjects would guess randomly for

that string. This two-parameter version of HTT should do better than the one-parameter version at predicting p(3Es)

In terms of PMT, it is plausible that p is not constant across subjects. Dienes (1992) assumed that subjects did not differ in their p's, but if different subjects have different learning rates, different noisy encodings of a stimulus at any one learning episode, and different starting weights, the same architecture (such as any of those in Dienes, 1992) would give different p's for different subjects, before learning asymptotes. A distribution produced by subjects with different p's is characteristically different to a binomial by virtue of having greater variance. This is exactly what the data showed - more p(3Cs) and p(3Es) than predicted by a binomial; that is, a greater variance. Consider, for example, an overall p of 0.5 that resulted from some subjects having ps greater than .50 and some less than .50. Then p(3Cs) and p(3Es) could be .50 each if a given subject had a p of either 0 or 1.

Different models of artificial grammar learning make different predictions about the behaviour of the parameters in two-parameter PMT. According to the connectionist models of Dienes (1992), subjects may start out with different p's, perhaps very discrepant ones if subjects start with very discrepant differences in weights, but the p's should converge with training - and at asymptote the variance in the p's should go to zero. On the other hand, Druhan and Mathews (1989) proposed a model of artificial grammar learning based on a classifier system, in which different subjects end up with different knowledge bases even after asymptotic learning. A prediction of the connectionist models but not the classifier system model is that one-parameter PMT should fit the data better and better as learning proceeds, and ultimately fit it perfectly, even for those strings for which the p remains very low. Experiments 1 and 3 both showed an increasingly good fit with training. Future research could determine whether the improvement continues with more extensive training.

Even if two-parameter HTT could account for subjects' consistency, it still in principle could not predict one finding from the data: that p(2Cs, 1E) was generally greater than p(1C, 2Es) given that $p > .60$. Inconsistent responding can only come from subjects who purely guess, and random guessing predicts that p(2Cs, 1E) should equal p(1C, 2Es). Augmenting two-parameter HTT with assumptions about subjects' memory for the previous response still could not explain the difference between p(2Cs, 1E) and p(1C, 2Es). The problem might be with the assumption that when subjects do not know the answer they guess randomly; indeed, in the No Exposure Group of Experiment 1, when subjects had not learnt anything, they did not respond randomly. A further extension of HTT would be to take the behaviour of subjects in a no training group as a model of subjects' guessing behaviour and then assume that the only influence of learning is to add correctly or incorrectly known items; i.e. the ratio of p(2Cs, 1E) to p(1C, 2Es) should not change with learning, only p(3Cs) and p(3Es)

would be incremented. However, in the No Exposure Group of Experiment 1, there were 90 cases classified correctly just twice and 92 cases classified correctly just once; i.e. p(2Cs, 1E) was virtually identical to p(1C, 2Es), if anything numerically smaller. Thus, it seems that it is the process of learning that produces the superiority of p(2Cs, 1E) over p(1C, 2Es); this is incompatible with any version of HTT. The superiority of p(2Cs, 1E) over p(1C, 2Es) is of consistent with two parameter PMT.

In summary, this paper has explored the way in which subjects apply their implicit knowledge of artificial grammars in terms of models suggested by Reber (1967) and Dienes (1992), and indicated how these models could be extended in future research.

References

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

Dienes, Z., & Berry, D. (1997). Implicit learning: below the subjective threshold. *Psychonomic Bulletin and Review, 4*, 3-23.

Dienes, Z., Broadbent, D. E., & Berry, D. C. (1991). Implicit and explicit knowledge bases in artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 17*, 875-882.

Dienes, Z., & Perner, J. (1996) Implicit knowledge in people and connectionist networks. In G. Underwood (Ed.), *Implicit cognition* (pp. 227-256). Oxford: Oxford University Press.

Druhan, B., & Mathews, R. (1989). THIYOS: A classifier system model of implicit knowledge of artificial grammars. *Proceedings of the Eleventh Annual Conference of the Cognitive Science Society*. New York: Erlbaum.

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

Reber, A.S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behaviour, 6*, 855-863.

Reber, A.S. (1989). Implicit learning and tactic knowledge. *Journal of Experimental Psychology: General, 118*, 219-235.

Reber, A. S. & Allen, R. (1978). Analogy and abstraction strategies in synthetic grammar learning: A functionalist interpretation. *Cognition, 6*, 189-221.

Reber, A. S., & Millward, R. B. (1968). Event observation in probability learning. *Journal of Experimental Psychology, 77*, 317-327.

Stephens, D. W., & Krebs, J. R. (1986). *Foraging theory*. Princeton, NJ: Princeton University Press.

Appendix 1. *Results for the No Exposure Group*
NB The critical value of χ^2 and the $p < .05$ level in each case is 7.81 (df=3).

Table 1.1. *Results for the No Exposure Group*

grammatical	3	2	1	0	P/p/X	χ^2
MTTTVT	19	3	7	16	.52	
pred. PMT	6.27	17.47	16.23	5.02	.52	67.10
pred. HTT	7.09	16.25	16.25	5.42	.04	56.73
MTVRXV	24	4	7	10	.64	
pred. PMT	12.04	19.94	11	2.03	.64	57.37
pred. HTT	17	12	12	4	.28	19.30
VXRRR	9	10	4	22	.38	
pred. PMT	2.43	11.99	19.75	10.84	.38	42.14
pred. HTT	4.25	12.75	12.75	15.25	.24	14.89
VXTTV	18	8	7	12	.57	
pred. PMT	8.35	18.87	14.21	3.57	.57	40.98
pred. HTT	11.16	14.5	14.5	4.83	.14	21.63
MVRXR	18	7	10	10	.58	
pred. PMT	8.68	19.03	13.91	3.39	.58	31.60
pred. HTT	11.75	14.25	14.25	4.75	.16	14.08
MTTVRX	20	13	6	6	.68	
pred. PMT	14.24	19.97	9.33	1.45	.68	20.23
pred. HTT	19.91	10.75	10.75	3.58	.36	4.21
non-grammatical	3	2	1	0	P/p/X	χ^2
MMVRX	20	8	9	8	.63	
pred. PMT	11.23	19.82	11.66	2.29	.63	28.74
pred. HTT	15.84	12.5	12.5	4.17	.26	7.21
VVXRM	18	12	7	8	.63	
pred. PMT	11.23	19.82	11.66	2.29	.63	23.27
pred. HTT	15.84	12.5	12.5	4.17	.26	6.25
TVTTXV	16	6	8	15	.50	
pred. PMT	5.75	17	16.75	5.5	.50	46.37
pred. HTT	5.91	16.75	16.75	5.58	.00	44.60
XVRXRR	11	13	9	12	.50	
pred. PMT	5.75	17	16.75	5.5	.50	17.00
pred. HTT	5.91	16.75	16.75	5.58	.01	16.20
MTRVRX	11	2	11	21	.36	
pred. PMT	2.03	11	19.94	12.04	.36	57.68
pred. HTT	4	12	12	17	.29	21.61
VXRVM	16	4	7	18	.47	
pred. PMT	4.58	15.68	17.92	6.83	.47	62.10
pred. HTT	5.25	15.75	15.75	8.25	.07	47.16

Table 1.2. *Results for the Single Exposure Group*

grammatical	3	2	1	0	P/p/X	χ^2
MTTTVT	24	5	2	2	.85	
pred. PMT	20.16	10.8	1.93	0.12	.85	33.30
pred. HTT	24.25	3.75	3.75	1.25	.70	1.69
MTVRXV	21	8	1	3	.81	
pred. PMT	17.41	12.4	2.95	0.23	.81	36.95
pred. HTT	21.91	4.75	4.75	1.58	.62	6.50
VXRRR	14	10	3	6	.66	
pred. PMT	9.34	14.66	7.67	1.34	.66	22.86
pred. HTT	13.16	8.5	8.5	2.83	.31	7.43
VXTTV	11	13	6	3	.66	
pred. PMT	9.34	14.66	7.67	1.34	.66	2.90
pred. HTT	13.16	8.5	8.5	2.83	.31	3.48
MVRXR	9	8	8	7	.52	
pred. PMT	4.51	12.74	11.99	3.76	.52	10.35
pred. HTT	5	12	12	4	.03	8.12
MTTVRX	10	5	14	4	.55	
pred. PMT	5.36	13.39	11.16	3.1	.55	10.26
pred. HTT	6.75	11.25	11.25	3.75	.09	5.73
nongrammatical	3	2	1	0	P/p/X	χ^2
MMVRX	22	6	2	3	.81	
pred. PMT	17.41	12.4	2.95	0.23	.81	38.18
pred. HTT	21.91	4.75	4.75	1.58	.62	3.20
VVXRM	16	11	5	1	.76	
pred. PMT	14.35	13.77	4.41	0.47	.76	1.42
pred. HTT	19	6	6	2	.52	5.31
TVTTXV	14	11	5	3	.70	
pred. PMT	11.17	14.57	6.34	0.92	.70	6.58
pred. HTT	15.5	7.5	7.5	2.5	.39	2.71
XVRXRR	7	7	12	7	.47	
pred. PMT	3.53	11.72	12.97	4.79	.47	6.40
pred. HTT	3.92	11.75	11.75	5.59	.05	4.70
MTRVRX	6	11	9	7	.49	
pred. PMT	4	12.25	12.5	4.25	.49	3.89
pred. HTT	4.08	12.25	12.25	4.41	.01	3.41
VXRVM	8	4	12	9	.44	
pred. PMT	2.9	10.86	13.58	5.66	.44	15.46
pred. HTT	3.67	11	11	7.34	.11	10.03

Table 1.3. *Results for the Double Exposure Group*

grammatical	3	2	1	0	P	χ^2
MTTTVT	30	16	2	1	.84	
pred. PMT	29.4	16.37	3.04	0.19	.84	3.83
pred. HTT	35.59	5.75	5.75	1.92	.69	22.04
MTVRXV	35	7	5	2	.84	
pred. PMT	29.4	16.37	3.04	0.19	.84	24.94
pred. HTT	35.59	5.75	5.75	1.92	.69	0.38
VXRRR	27	15	4	3	.78	
pred. PMT	23.46	19.59	5.45	0.5	.78	14.50
pred. HTT	30.34	8	8	2.67	.56	8.53
VXTTV	10	12	18	9	.49	
pred. PMT	5.76	17.99	18.74	6.51	.49	6.10
pred. HTT	6	18	18	7	.02	5.24
MVRXR	12	13	14	10	.52	
pred. PMT	6.77	18.98	17.73	5.52	.52	10.35
pred. HTT	7.59	17.75	17.75	5.92	.03	7.44
MTTVRX	11	12	14	12	.48	
pred. PMT	5.52	17.73	18.98	6.77	.48	12.64
pred. HTT	5.92	17.75	17.75	7.59	.03	9.58
nongrammatical	3	2	1	0	P	χ^2
MMVRX	38	7	4	0	.90	
pred. PMT	35.49	12.09	1.37	0.05	.90	7.42
pred. HTT	40.25	3.75	3.75	1.25	.80	4.21
VVXRM	30	11	7	1	.81	
pred. PMT	25.99	18.35	4.32	0.34	.81	6.51
pred. HTT	32.66	7	7	2.33	.62	3.26
TVTTXV	25	13	8	3	.74	
pred. PMT	19.98	20.89	7.28	0.85	.74	9.75
pred. HTT	26.84	9.5	9.5	3.17	.48	1.66
XVRXRR	19	17	10	3	.69	
pred. PMT	15.9	21.72	9.89	1.5	.69	3.13
pred. HTT	22.16	11.5	11.5	3.83	.37	3.46
MTRVRX	11	13	16	9	.51	
pred. PMT	6.51	18.74	17.99	5.76	.51	6.90
pred. HTT	7	18	18	6	.02	5.40
VXRVM	8	12	17	12	.44	
pred. PMT	4.24	16.03	20.23	8.51	.44	6.29
pred. HTT	5.42	16.25	16.25	11.09	.12	2.45

Appendix 2. *Results for Experiment 2*

NB The critical value of χ^2 and the $p < .05$ level in each case is 7.81 (df=3).

grammatical	3	2	1	0	P/p/X	χ^2
MTTTZT	8	7	9	6	.52	
pred. PMT	4.27	11.73	10.73	3.27	.52	7.72
pred. HTT	4.91	10.75	10.75	3.58	.04	5.17
MTZRXZ	17	8	3	2	.78	
pred. PMT	14.12	12.1	3.46	0.33	.78	10.49
pred. HTT	18.34	5	5	1.67	.56	2.76
ZXRRR	14	3	7	6	.61	
pred. PMT	6.85	13.07	8.32	1.76	.61	25.65
pred. HTT	9.59	8.75	8.75	2.92	.22	9.40
ZXTTZ	15	10	3	2	.76	
pred. PMT	12.94	12.56	4.06	0.44	.76	6.66
pred. HTT	17.16	5.5	5.5	1.83	.51	5.11
MZRXR	15	5	4	6	.66	
pred. PMT	8.51	13.33	6.96	1.21	.66	30.38
pred. HTT	11.92	7.75	7.75	2.58	.31	8.13
MTTZRX	10	11	8	1	.67	
pred. PMT	8.89	13.33	6.67	1.11	.67	0.82
pred. HTT	12.5	7.5	7.5	2.5	.33	3.07
non-grammatical	3	2	1	0	P/p/X	χ^2
MMZRX	3	10	12	5	.46	
pred. PMT	2.84	10.17	12.15	4.84	.46	0.02
pred. HTT	3.42	10.25	10.25	6.09	.09	0.55
ZZXRM	8	6	11	5	.52	
pred. PMT	4.27	11.73	10.73	3.27	.52	6.98
pred. HTT	4.91	10.75	10.75	3.58	.04	4.61
TZTTXZ	14	10	4	2	.73	
pred. PMT	11.83	12.91	4.7	0.57	.73	4.75
pred. HTT	16	6	6	2	.47	3.58
XZRXRR	5	5	7	13	.36	
pred. PMT	1.35	7.34	13.29	8.03	.36	16.67
pred. HTT	2.67	8	8	11.34	.29	3.53
MTRZRX	8	5	9	8	.48	
pred. PMT	3.27	10.73	11.73	4.27	.48	13.80
pred. HTT	3.58	10.75	10.75	4.91	.04	13.56
ZXRZM	10	13	2	5	.64	
pred. PMT	8.03	13.29	7.34	1.35	.64	14.24
pred. HTT	11.34	8	8	2.67	.29	9.82

Appendix 3. *Results for Experiment 3*

NB The critical value of χ^2 and the $p < .05$ level in each case is 7.81 (df=3).

Table 3.1. *Results for the Single Exposure Group*

grammatical	3	2	1	0	P/p/X	χ^2
MTTTVT	25	8	9	10	.64	
pred. PMT	13.52	22.88	13	2.6	.64	41.72
pred. HTT	19.34	14	14	4.67	.28	12.10
MTVRXV	33	12	5	2	.82	
pred. PMT	28.6	18.72	4.16	0.52	.82	7.47
pred. HTT	35.66	7	7	2.33	.64	4.39
VXRRR	22	16	7	7	.67	
pred. PMT	15.6	22.88	11.44	1.56	.67	25.39
pred. HTT	22.25	12.75	12.75	4.25	.35	5.20
VXTTV	20	19	6	7	.67	
pred. PMT	15.6	22.88	11.44	2.08	.67	16.12
pred. HTT	21.66	13	13	4.33	.33	8.31
MVRXR	13	16	12	11	.53	
pred. PMT	7.8	20.8	18.2	5.2	.53	13.16
pred. HTT	9.41	18.25	18.25	6.08	.06	7.77
MTTVRX	11	13	16	12	.48	
pred. PMT	5.72	18.72	20.28	7.28	.48	10.58
pred. HTT	6.25	18.75	18.75	8.25	0.04	7.48
nongrammatical	3	2	1	0	P/p/X	χ^2
MMVRX	36	9	5	2	.84	
pred. PMT	30.68	17.68	3.12	0	.84	6.32
pred. HTT	37.41	6.25	6.25	2.08	.68	1.52
VVXRM	31	13	6	2	.80	
pred. PMT	26.52	19.76	4.68	0.52	.80	7.65
pred. HTT	33.91	7.75	7.75	2.58	.60	4.33
TVTTXV	19	17	13	3	.67	
pred. PMT	15.6	22.88	11.44	2.08	.67	2.87
pred. HTT	21.66	13	13	4.33	.33	1.97
XVRXRR	14	11	13	14	.49	
pred. PMT	6.24	19.24	19.76	6.76	.49	23.25
pred. HTT	6.42	19.25	19.25	7,09	.01	21.25
MTRVRX	6	12	18	16	.38	
pred. PMT	3.12	14.04	22.88	11.96	.38	5.36
pred. HTT	6.5	19.5	19.5	6.5	0	16.92
VXRVM	7	5	16	24	.30	
pred. PMT	1.56	9.88	22.88	17.68	.30	25.71
pred. HTT	3.92	11.75	11,75	24,59	.40	7.849

Table 3.2. *Results for the Double Exposure Group*

grammatical	3	2	1	0	P/p/X	χ^2
MTTTVT	27	12	4	4	.77	
pred. PMT	21.62	19.27	5.64	0.47	.77	26.51
pred. HTT	28.34	8	8	2.67	.55	3.88
MTVRXV	38	7	2	0	.92	
pred. PMT	36.66	9.4	0.94	0	.92	0
pred. HTT	40.59	2.75	2.75	0.92	.84	2.73
VXRRR	25	11	6	5	.73	
pred. PMT	18.33	20.21	7.52	0.94	.73	17.54
pred. HTT	24.84	9.5	9.5	3.17	.46	5.56
VXTTV	24	12	3	8	.70	
pred. PMT	16.45	20.68	8.93	1.41	.70	30.8
pred. HTT	22.5	10.5	10.5	3.5	.40	13.37
MVRXR	14	19	8	6	.62	
pred. PMT	11.28	20.68	12.22	2.35	.62	5.67
pred. HTT	16.09	13.25	13.25	4.42	.25	2.16
MTTVRX	15	13	11	8	.58	
pred. PMT	9.4	19.74	14.57	3.29	.58	6.74
pred. HTT	12.59	14.75	14.75	4.92	.16	5.1
nongrammatical	3	2	1	0	P/p/X	χ^2
MMVRX	41	4	2	0	.94	
pred. PMT	39.48	7.05	0.47	0	.94	6.36
pred. HTT	39.41	3.25	3.25	1.08	.82	1.80
VVXRM	33	10	4	0	.87	
pred. PMT	31.02	13.63	1.88	0	.87	3.48
pred. HTT	33.59	5.75	5.75	1.92	.74	5.6
TVTTXV	17	18	6	6	.66	
pred. PMT	13.63	20.68	10.81	1.88	.66	12.35
pred. HTT	16.09	13.25	13.25	4.42	.25	6.29
XVRXRR	11	6	19	11	.45	
pred. PMT	4.23	15.98	19.27	7.52	.45	18.68
pred. HTT	5.33	16	16	9.66	.09	13.03
MTRVRX	6	8	20	13	.38	
pred. PMT	2.82	12.69	20.68	10.81	.38	5.79
pred. HTT	4.5	13.5	13.5	15.5	0.23	6.27
VXRVM	5	12	19	11	.41	
pred. PMT	3.29	14.1	20.21	9.4	.41	1.55
pred. HTT	4.83	14.5	14.5	13.16	0.18	2.19

Table 3.3. *Results for the Triple Exposure Group*

grammatical	3	2	1	0	P/p/X	χ^2
MTTTVT	22	7	5	4	.75	
pred. PMT	15.58	15.96	5.32	0.76	.75	13.81
pred. HTT	21.09	7.25	7.25	2.42	.49	6.08
MTVRXV	28	9	1	0	.90	
pred. PMT	28.12	9.12	1.14	0	.90	0
pred. HTT	31.59	2.75	2.75	0.92	.81	0.94
VXRRR	26	5	6	1	.82	
pred. PMT	21.28	13.68	3.04	0.38	.82	1.01
pred. HTT	26.34	5	5	1.67	.65	8.66
VXTTV	20	10	5	3	.75	
pred. PMT	15.58	15.96	5.32	0.76	.75	6.6
pred. HTT	21.09	7.25	7.25	2.42	.49	2.38
MVRXR	22	8	6	2	.77	
pred. PMT	17.48	15.58	4.56	0.38	.77	6.91
pred. HTT	22.84	6.5	6.5	2.17	.54	4.16
MTTVRX	11	10	12	5	.57	
pred. PMT	7.22	15.96	12.16	3.04	.57	1.26
pred. HTT	9.41	12.25	12.25	4.08	.14	2.44
nongrammatical	3	2	1	0	P/p/X	χ^2
MMVRX	34	3	1	0	.96	
pred. PMT	33.06	4.56	0.38	0	.96	1.57
pred. HTT	35.09	1.25	1.25	0.42	.91	2.95
VVXRM	34	2	0	2	.93	
pred. PMT	30.4	6.84	0.38	0	.93	4.23
pred. HTT	33.34	2	2	0.67	.86	4.65
TVTTXV	20	8	9	1	.75	
pred. PMT	15.58	15.96	5.32	0.76	.75	7.85
pred. HTT	21.09	7.25	7.25	2.42	.49	1.39
XVRXRR	8	10	6	14	.44	
pred. PMT	3.04	12.16	15.58	6.84	.44	21.86
pred. HTT	4.17	12.5	12.5	8.84	.12	10.41
MTRVRX	2	10	10	16	.32	
pred. PMT	1.14	7.6	16.72	12.16	.32	5.32
pred. HTT	3	9	9	17	.37	.61
VXRVM	5	4	15	14	.33	
pred. PMT	1.52	8.36	16.72	11.4	.33	11.01
pred. HTT	3.17	9.5	9.5	15.84	.33	7.64

Received April, 1997