THE QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY, 1998, *51B* (3), 235–270

Perceptual Learning and Free Classification

A. J. Wills and I. P. L. McLaren
University of Cambridge, Cambridge, U.K.

Two experiments are reported that investigate the effects of stimulus preexposure on discrimination performance in a free classification task, using adult humans as subjects. In free classification subjects are asked to put stimuli into gruops in any way that seems reasonable or sensible to them. Experiment 1 shows that the effect of preexposure is contingent on stimulus structure. Experiment 1b is the first demonstration of a retardation in learning as a consequence of simple preexposure in adult human subjects (previous demonstrations have relied on incidental or masked preexposure). Experiment 2 further supports the conclusions of Experiment 1 and extends them with the demonstration that stimulus similarity is a crucial factor. Taken together, these experiments rule out a class of attention-based explanations of the phenomena reported here. The experiments also provide novel information about the effects of preexposure. Preexposure can change the actual classifications subjects form in addition to altering the rate at which they are formed. Implications of these results for current theories of category formation and perceptual learning are considered.

In Bruner, Goodnow, and Austin's (1956) classic investigations of category learning there are three main methodological conventions—the use of perceptually simple stimuli, the presence of immediate and accurate membership information for each example seen, and the use of categories defined by a simple logical rule. Even at that time the last of these conventions was not universally adopted, and it has subsequently fallen out of favour. The difficulty of defining "natural" categories in terms of singly necessary and jointly sufficient conditions highlighted by Rosch (Rosch, 1973; Rosch & Mervis, 1975) and by Wittgenstein (1958) led many investigators to direct their attention to stochastically defined categories in which no single cue is entirely valid (e.g. Medin & Schaffer, 1978). The remaining two conventions, however, have gone largely unchallenged over the intervening decades.

We define "perceptually simple stimuli" as those for which a naïve observer, given the full set, could easily determine the attributes relevant for classification and could easily discriminate between the different values that those attributes could take. For example, one would have little difficulty in determining that the single geometric shapes used by

Requests for reprints should be sent to A.J. Wills, The Psychological Laboratory, University of Cambridge, Downing Street, Cambridge, CB2 3EB, U.K. E-mail: ajw43@cus.cam.ac.uk

This research was supported by a grant from the BBSRC to I.P.L. McLaren and a BBSRC studentship held by A.J. Wills. The authors would like to thank N.J. Mackintosh and two anonymous reviewers for their helpful comments.

Shepard, Hovland, and Jenkins (1961, Exp. II and III, representation A) were either triangular or square, large or small, and black or white. Given this definition, the vast majority of the studies considered in three recent reviews of categorization research (Estes, 1994; Shanks, 1995; Smith & Medin, 1981) use simple stimuli. Taking the same three reviews, it is clear that a similar conclusion can be reached for immediate, example-specific, category label information. Indeed, one might be forgiven for thinking that "category-learning experiment" was synonymous with "two category, trial and error with feedback". Estes acknowledges this in his review: "The question of how a learner in a natural environment selects the categories to be formed from among all those that could be formed is beyond the scope of the theories developed or reviewed in this book" (Estes, 1994, p. 242).

Clearly, it can be beneficial to simplify the categorization tasks administered to subjects for experimental purposes, yet in attempting to make a difficult problem more tractable we may be in danger of obscuring something important. It seems likely that learning to categorize involves more than learning example–label correspondences. Equally, it is not clear that in all important situations we would or could know a priori the appropriate number of groups to use, or that our environment would act as an omnipresent, totally reliable, item-by-item teacher. The number-of-groups constraint doesn't hold for a variety of situations—learning about wild birds, grouping a collection of books or recordings, and so forth—and it seems unlikely that the total-feedback constraint would hold even in the most scholastic forms of learning. How do we learn to categorize in these circumstances? Can categories be developed as a consequence of experience with the appropriate stimuli in the absence of explicit instruction?

Our interest in category development was prompted by consideration of the role that perceptual learning might play in this process. The basic perceptual learning effect is that simple preexposure to stimuli can enhance a subject's ability to discriminate between them. The effect has been found in rats (Gibson & Walk, 1956; Hall, 1979) and in undergraduates (Goss, 1953). We also know that, in undergraduates, learning to categorize stimuli such as chequerboard patterns into two classes (each defined by a prototype) leads to better discrimination both between and within categories (McLaren, Leevers, & Mackintosh, 1994; McLaren, 1997). In pigeons, preexposure to a chequerboard pattern leads to faster learning of a discrimination between two distortions of that pattern (Aitken, Bennett, McLaren, & Mackintosh, 1996).

An experiment in the McLaren et al. (1994) paper was the starting-point for the current investigation. In the experiment of interest, a different pair of randomly generated chequerboard patterns was created for each subject. Being randomly generated, the patterns were generally somewhat similar, sharing on average 50% of squares. The two created patterns served as prototypes; the patterns actually presented to subjects were distortions of these prototypes created by changing a proportion of the squares, at random, from black to white or vice versa. Patterns were presented one at a time and subjects learned to categorize them, with feedback, into one of two groups. These groups corresponded to the sets of patterns created from the two prototypes.

After learning to categorize to criterion, subjects had to learn, simultaneously, four 2-item discriminations. The pairs of stimuli used in these four discriminations were (a) the two prototypes of the trained categories, (b) two new exemplars taken from one of the trained categories, (c) the prototypes of the previous subject's categories, and (d) two exemplars from one of the categories of the previuos subject. Pairs (c) and (d) should be relatively unfamiliar to subjects because each subject was trained to categorize examples from a different pair of prototypes. Subjects made fewer errors on the discriminations involving stimuli drawn from familiar categories than on discriminations involving stimuli drawn from unfamilar categories, supporting the claim that discrimination between and within familiar categories was better than that for novel categories.

This pattern of results can be explained by assuming that experience in categorization can lead to perceptual learning—that is, exposure to the stimuli during categorization leads to them being easier to discriminate per se, over and above any effect of the category information received. Such a process (whose proposed mechanism will be described later) could, no doubt, help "fine-tune" categorization decisions, but we do not know the extent to which perceptual learning resulting from familiarization with category exemplars might influence category formation itself. We can ask whether preexposure has any effect on category formation and, if it does, whether it simply accelerates the process of category formation that would have occurred in any case or if it can also modify the final solution adopted in response to the categorization problem.

The distinction between the rate of category formation and its final solution is important in situations where more than one mapping from stimulus to response is allowed. For example, given 16 wines, you could classify them as red, white, or rosé. Alternatively, you could group them according to whether they were Chenin Blanc, Muscat, Carbernet Sauvignon, or Merlot. A third option is to classify them on the basis of whether they might best accompany meat, poultry, fish, or dessert. The question of which classification you use is distinct from the question of how accurately you apply the classification to the sample of wines—for example, the extent to which you confuse Carbernet Sauvignon and Merlot. Would preexposure to wines simply lead to a more accurate classification along the same lines that would have occurred without preexposure, or might it also affect the classification system used? It is difficult to distinguish between the type of classification used and the adequacy with which it is employed in a standard category-learning procedure (such as two-category guess and correct-with-feedback) because only one stimulus-to-response mapping is allowed. We need a procedure that allows us to study category formation rather than just the acquisition of example–label correspondences.

Sequential free classification is just such a procedure. In free classification studies, subjects are asked to put stimuli into groups in any way that seems sensible or reasonable to them. In sequential free classification, decisions about stimuli are made on an item-by-item basis, and previous decisions cannot be referred back to or changed. This procedure has previously been used by Evans and his associates (Bersted, Brown, & Evans, 1969; Evans & Arnoult, 1967). Conclusions that can be drawn from his studies are limited, but they do demonstrate that, at least for the histoform stimuli used, classifications produced by some subjects are consistent with the underlying stimulus structure. In addition, Bersted et al. (1969) showed that when subjects are unconstrained as to the number of categories they can form, they tend to create more categories than experimentally defined.

The phrases "consistent with underlying stimulus structure" and "more categories than experimentally defined" may require some explanation. In these studies, the experimenters have created examples that, due to the way they are constructed, are considered by the experimenters to have come from three coherent groups. Subjects' reponses are consistent with this underlying stimulus structure to the extent that the individual groups they use in their classifications contain examples that come predominantly from just one of the experimenter's groups. A subject can be "consistent" whilst using more categories than experimentally defined if not all examples from a single experimenter's group fall into a single group used by the subject. Bersted et al. (1969) present some evidence that subjects are subdividing the experimenters' categories into high- and low- typicality sub-groups. How one may properly define and assess this form of "consistency" is considered in a later section.

The experiments reported in this paper are, as far as we are aware, the first to investigate the relationship between perceptual learning and free classification. As such, our investigations should be taken for what they are—interesting, initial explorations of the area, rather than a closed set of experiments providing a definitive answer to a single question. In our first experiment we set out to develop a free-classification task that would allow us to study the effects of stimulus preexposure on category formation. Experiment 1a reports our attempt to extrapolate from the McLaren et al. (1994) type of study to the free classification paradigm using chequerboard stimuli structured around prototype-defined categories. We predicted that preexposing subjects to the type of stimuli they would later be asked to free-classify would increase the "consistency" of their classifications. We decided to investigate both two- and four-category problems to increase the generality of any conclusions that we might come to.

EXPERIMENT 1

Dependent Measures

Previous studies of preexposure have used some measure of response accuracy to index its effects—for example, the number of trials required to reach a criterial level of performance (Gibson & Walk, 1956) or the overall percentage of correct responses (McLaren et al., 1994). *Percentage correct* and *trials to criterion* are both measures of the extent to which the subjects are behaving as the experimenter wants or expects them to. With animal experiments we typically tie our intentions for the subject to some event of primary motivational significance for it. With undergraduates, the motivation for responding correctly is probably more subtle.

In free classification, more than one stimulus to response mapping is allowed. Therefore it is not possible to assess accuracy in the same way because no one response can be considered to be individually right or wrong. Any single response is neither what the experimenters expect for "good" performance or what they do not expect. However, the extent to which a set of responses follows our expectations can be assessed. In the previous section we considered this informally by saying that if subjects are following the experimenter-defined stimulus structure then we would expect each of their groupings to consist predominantly of examples from just one experimenter-defined group. We also indicated that the reverse need not hold—if a subject divides a single experimenter's group into a number of sub-groups this can still be considered a good classification as long as each sub-group predominantly contains examples from a single experimenter's group. In a sense, we are considering the experimenter's classification as the coarsest allowable perfect classification. The assessment of the "goodness" of subjects' responses is lowered if they reduce different experimenter-defined groups to a single group, but not if they sub-divide them. This measure of goodness is clearly not identical to accuracy as it is more normally defined so, in an attempt to avoid confusion, we refer to this quality as "consistency". The term derives from the fact we are assessing the extent to which a subject's classifications are consistent with the experimenter's classifications.

This informal consideration of what is meant by consistency is necessary, but it is not sufficient. How does one get from this general notion to a single variable that captures the "goodness" of free-classification decisions without being artifactually affected by the number of groups the experimenter and the subject use? The statistic we adopted was based on Cramér's phi (Cramér, 1946), which is designed to allow measurement of the association (coprediction) between two categorical variables. Calculation of Cramér's phi (Φ_c) begins by computing the chi-square statistic (without correction for continuity) from a c by r contingency table of the two variables (see e.g. Howell, 1992, p. 147–148). In our case c will be the number of experimenter-defined categories presented and r the number of subject-defined groups containing at least one stimulus. If we define k as the smaller of r and c, and N as the number of stimuli grouped, then Cramér's phi is given by:

$$\phi_c = \sqrt{\frac{\chi^2}{N(k-1)}} \qquad 1$$

Phi has the desirable characteristic of varying from 0 (no association) to 1 (maximum association) for any r and c, but as it stands this statistic has a serious drawback. Its expected value will vary with the size of the contingency table—the larger the values of r and c the larger the expected value of phi. This makes it invalid to compare between subjects using different numbers of groups and categories. This is precisely what we wanted to do in the experiments reported below, so this limitation was unacceptable. Our solution was to use phi adjusted for the value expected by chance and then scaled such that 1 represents perfect association and 0 the chance level. Formally, ϕ_{adj} was given by:

$$\phi_{adj} = \frac{\phi_c - \phi_{chance}}{1 - \phi_{chance}} \qquad 2$$

ϕ_{chance} is, to a first approximation, given by substituting $(r-1).(c-1)$ for χ^2 in Equation 1. This approximation is based on an assumption that the chance expectation can be derived from the mean value of χ^2, which is given by its degrees of freedom (see Howell, 1992). Such an approach, however, takes no account of the effect that the square-root transform has on the chi-square distribution, and a more accurate approximation makes use of the fact that $\sqrt{\chi^2}$ is approximately normally distributed with mean

$$\sqrt{\frac{2df - 1}{2}} \qquad 3$$

(see Rosenthal & Rosnow, 1984, p. 457). This gives

$$\phi_{\text{chance}} = \sqrt{\frac{2(r - 1)(c - 1) - 1}{2N(k - 1)}} \qquad 4$$

In order to assess the usefulness of this approximation we used numerical methods, running a separate "Monte Carlo" simulation (1×10^6 iterations) of random responding to an equal number of examples from each experimenter-defined category, for each size of contingency table that could potentially arise in the forthcoming experiments ($2 \times 2, \ldots, 10$, and $4 \times 2, \ldots, 10$). We found a good fit to the approximation given here for ϕ_{chance}, but an even better fit could be obtained by replacing N in the denominator with $N - 1$, giving

$$\phi_{\text{chance}} = \sqrt{\frac{2(r - 1)(c - 1) - 1}{2(N - 1)(k - 1)}} \qquad 5$$

which is the approximation used in this paper. We believe that this small modification stems from the fact that we are dealing with contingency tables in which the marginal row totals are always equal (because the same number of examples from each category are presented) and in which all cells are constrained to contain at least one count.

Given this formula for ϕ_{chance}, ϕ_{adj} is always 1 for perfect association (coprediction) and 0 for the degree of association expected by chance, independent of the size of the contingency table. ϕ_{adj} gives us a measure of the consistency (as previously defined) of a set of a single subjects' classification decisions, with respect to the category structure defined by the experimenter. The measurement that it gives is independent of the number of groups the subject and experimenter use.

To illustrate this latter property, consider two different subjects in an experiment with two categories (A and B). Subject 1 uses two groups (1 and 2) and always responds "2" to Category A and "1" to Category B. Her ϕ_{adj} will be 1. Subject 2 uses four groups (3, 4, 5, and 6). He always responds "6" or "3" to Category A and "5" or "4" to Category B. His ϕ_{adj} will also be 1. This basic principle also holds for non-perfect classifications. Of course, ϕ_{adj} can not compensate for the fact that it might be harder to keep four groups in mind than two and so the number of groups a subject uses may well affect the subject's consistency score. However, the reduction one might expect is not an artifact of the statistic—it represents a real difference in the consequences of using different classification systems. A similar point can be made about changes in the number of categories the experimenter uses in constructing the stimuli. Raising this number may increase the difficulty of the task subjects' face and so may reduce their consistency score. However, this will be as a result of their making more errors, and not due to a simple artifact of increasing the size of the contingency table. Whatever the number of experimenter-defined categories, perfect responding will produce a ϕ_{adj} of 1, and chance responding will produce ϕ_{adj} scores that average to zero.

As our measure of consistency is independent of the number of groups a subject uses, we can employ the number of groups used as a second orthogonal dependent variable. Our first experiment examined the effect of preexposure on the consistency of subjects' classifications when faced with a two-category problem or with a four-category problem and its effect on the number of groups they used. Free classification was divided, by pauses, into a series of 15 equal blocks, so that changes in consistency with increasing classification experience could be assessed.

Method

Subjects and Apparatus

The subjects were 48 adults, aged between 18 and 30, who were paid for their participation. Most were graduate or undergraduate students at Cambridge University. Subjects were tested individually in one of two quiet experimental cubicles. Each cubicle contained an Acorn microcomputer connected to a 14-in. colour monitor. The two computers were not identical—one was an Acorn A5000/ AKF50 and the other was an Acorn Risc PC600/ AKF60—but these two machines were very similar, and the program ran identically on both without modification. Subjects sat about 1 m from the screen, which was approximately at eye level.

Stimuli

Each stimulus was a 16×16 array of black and white squares. These "chequerboards" measured 2.5 cm on a side and were presented in the centre of the screen against a mid-grey background. For each subject a new *master pattern* was generated, and four prototype patterns were produced from it. The master pattern was a chequerboard of 128 white squares and 128 black squares, randomly placed. Each prototype differed from the master pattern by exactly 64 squares, a completely different 64 being selected for each prototype. The process of creating the four prototype patterns was random within these constraints.[1] Squares were changed by reversing their shade (black to white or vice versa). Figure 1 shows four example prototype patterns. The examples that were actually presented to subjects were created from the prototypes by subjecting each square to a small independent chance ($p = .05$) of reversing its colour.

Procedure

The experiment was of a 2×2 factorial between-subjects design; the two factors were preexposure (2 levels) and a number of categories (2 levels). Hence there were four between-subject conditions—preexposed two category, preexposed four category, non-preexposed two category and non-preexposed four category. The main difference between two-category and four-category conditions was in the stimuli presented. In the four-category conditions all four

[1] If random creation of four prototypes, each of which differs from a master pattern by exaclty 64 different squares, seems untenable, then consider the following example. Imagine a bag, which contains 64 chips with a "1" written on, 64 with a "2", 64 with a "3", and 64 with a "4". Shake the bag well, and then lay one chip on each square of the master pattern. Now, to construct the four prototype patterns, which we shall call 1, 2, 3, and 4, change the colour of squares containing a correspondingly numbered chip, leaving the colour of the remainder unaltered. The stimuli used in this experiment were created with the aid of a computer, but the process was isomorphic with the one described here.

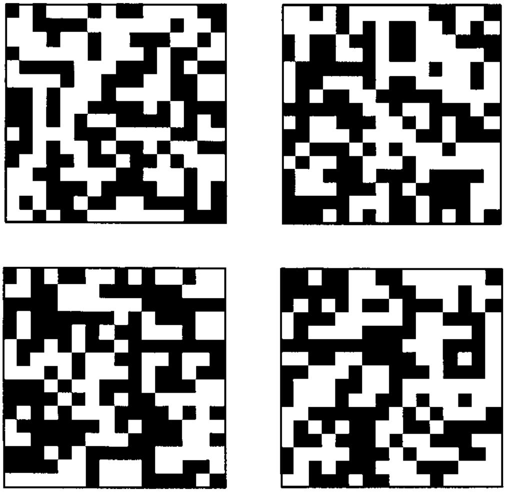

Fig. 1. Four examples of the type of chequerboard stimuli used in Experiment 1. Four prototype patterns created from the same master pattern are shown. The rectangle outlines enclosing the stimuli represent the beginning of the mid-grey background against which they were presented; they are not part of the stimulus itself and were not presented to subjects.

prototypes were used to create stimuli, whereas in the two-category conditions only two prototypes, randomly selected at the beginning of the experiment, were used.

Each condition has two phases. In the preexposed conditions Phase 1 was running recognition and Phase 2 was free classification (procedures given below). In the non-preexposed conditions, Phase 1 was an unrelated experiment of approximately the same duration as the running-recognition phase, and Phase 2 was again free classification.

Running-Recognition Phase. Subjects were given instructions on a printed sheet, and any questions were answered by reiterating the printed instructions. These instructions were as follows (paragraphing lost for compactness):

> In this experiment you will be shown a lot of items. Here is an example of one item on a grey background: [example chequerboard pattern]. The items will appear on the computer screen one at a time. Each item will be shown exactly twice. Your task is to decide, for each shape, whether it is the first or the second time you have seen it. If it is the first time, you should press [picture of the X key] on the keyboard. If it is the second time, you should press [picture of the > key] on the keyboard. You will have quite a bit of time to make each of these decisions. However, if you take too long, the computer will beep and tell you so. Try to avoid this happening. The computer will also beep if you press any key apart from the two shown above. If this happens please tell the computer your decision again, using one of the proper keys. After you have put about fifty items into groups the computer will stop. At this point, all the shapes will have been shown twice. This is the end of a ‘block’. There are [five or ten] blocks in this experiment, and each block uses a totally different set of shapes. At the end of a block it is a good idea to rest for a few seconds. When you are ready to start again, press Y on the keyboard.

The “X” and “>” keys are both on the bottom row of a standard computer keyboard. Immediately after the subject had responded to a pattern, the pattern disappeared and was replaced by the next one. Subjects were given a maximum of 5 sec to respond before being timed-out and asked to respond more quickly in future. After making 48 decisions, the computer signalled the end of a block by clearing the screen and requiring a key press. There were either five or ten blocks of trials. Five blocks were used in the preexposed-two-category condition and ten in the preexposed-four-category condition, thereby equalizing the amount of exposure given per category.

Within a block, 24 different chequerboards were each presented twice. All were examples created from a relevant prototype, with an equal number coming from each one. In the preexposed-four-category condition, six examples from each of the four prototypes were used, whereas in the preexposed-two-category condition 12 examples from each of two prototypes were used. A check was made to ensure that across all blocks any given pattern only occurred twice. If this was not the case then a whole new set of randomly generated examples was created. This was made possible by creating all patterns and checking them before presenting any to the subject (the subject did not observe this process).

The ordering of stimuli within a block was a little complex. First, all of the 24 different stimuli to be presented in a block were placed in a list in a random order. Second, each of the 48 trials in a block was randomly designated as a “no” trial (stimulus not presented before) or a “yes” trial (stimulus presented once before) with the constraints that the first three trials were no trials, the last was a yes trial, and there were an equal number of no and yes trials. Then two pointers were placed at the beginning of the stimulus order list (one for no trials and one for yes trials). Finally, for each trial in order the stimulus indicated by the appropriate pointer was inserted into a final list, and the pointer was moved to the next item in the order list (the example given in Figure 2 should make this stage somewhat clearer). This somewhat convoluted procedure was designed to reduce the effectiveness of position within a block as a cue for if a stimulus was being presented for the first or second time (at least for the range trial 4 to trial 47).

Free-Classification Phase. Subjects were given instructions on a printed sheet and questions were answered by reiterating the printed instructions. The sheet read as follows:

> This experiment involves putting things into groups. Here is an example of one item, on a grey background [example chequerboard pattern]. The items will appear on the computer screen one at a time. At first, all the items might seem to be very similar. However, they can actually be put

```
Trial 1          Trial 2          Trial 3          Trial 4 ...
Y                Y                Y                Y
6 2 3 1 (6)      6 2 3 1 (2)      6 2 3 1 (3)      6 2 3 1 (6)
N                  N                  N                  N

Sequence of yes and no trials : N, N, N, Y ...
Pattern presented to subject  : 6, 2, 3, 6 ...
```

FIG 2. An example of the method of trial ordering used in the running-recognition phase of all experiments. The four sub-figures titled “Trial 1”, “Trial 2” etc. are the first four trials of the running-recognition phase. The Y and N in bold type represent the pointer for “yes” trials and “no” trials, respectively. The list of four numbers are the first four chequerboards to be presented (they read 6, 2, 3, 1 rather than 1, 2, 3, 4 to illustrate that the patterns have been randomly ordered). The number in brackets is the pattern actually presented on that trial. The list of Ns and Ys in plain type is the sequence of yes and no trials. The sequence of yes and no trials determines for each trial which pointer indicates the stimulus to be presented and which pointer is ignored. The chosen pointer is moved one item along the list after stimulus selection and before the next trial. On trial 5, the Y pointer would be over the 2.

> into a number of groups and this is what you will be asked to do. When an item appears on the screen, you put it in a group by pressing a key on the keyboard. You can use any of the following keys: [picture of the number keys 1 to 9 plus 0 on the top row of a computer keyboard]. To start with, you will obviously be guessing. Do not worry about this. As you put more and more items into groups, you should get more idea about how they should be grouped. As you go along, you may think that there are more groups than you are currently using. Alternatively, you may decide that you are using too many groups and it is actually better to use less. If either of these things happen, feel free to use more or less keys as appropriate. You have quite a bit of time to put each object into a group. However, if you take too long, the computer will beep and tell you so. Try to avoid this happening. The computer will also beep if you press any key apart from the ones shown above. If this happens, please put the item into a group using one of the proper keys. After you have put about fifty items into groups the computer will stop. This is the end of a block, not the end of the experiment. At this point you can rest for a few seconds, but it is very important that you do not forget which keys you have been using. When you are ready to start again, press Y on the keyboard. There are fifteen blocks.

No explicit reference to the previous phase was made.

As soon as a subject had responded to a pattern, the pattern disappeared and was replaced by another. Subjects had a maximum of 5 sec to reach a decision about each pattern. After 48 patterns had been classified, the computer signalled the end of a block by clearing the screen and requiring a key press to continue.

Subjects classified 15 blocks of patterns. In the four-category conditions, 12 examples from each of the four prototypes were presented in a random order. In the two-category conditions, 24 examples from each of the two selected prototypes were presented, again in random order. If the subjects had just completed a running-recognition phase, then the prototypes used to construct the examples were the same as in that phase.

As a consequence of our design, subjects in the two-category conditions saw twice as many examples per category overall as did subjects in the four-category conditions. However, it also meant that subjects classified the same number of stimuli in one block of the four-category conditions as the number in one block of the two-category conditions. As we wished to compare the consistency scores of subjects in the two- and four-category conditions it seemed important that the experience they had within a block was otherwise equivalent.

Results

Separate one-sample t tests ($\mu = 0$), were run on subjects' adjusted Cramér's phi scores on the last block of each of the four conditions. In all conditions, scores were significantly greater than zero, $t(11) = 25$, 9.3, 18, and 7.6 for the two-category–non-preexposed, four-category–non-preexposed, two-category–preexposed, and four-category–preexposed conditions, respectively. In other words, subjects in all four conditions were, as a group, making classifications reliably more consistent with the underlying stimulus structure than would be expected by chance. However, closer inspection of individual subject data revealed some subjects whose scores across 15 blocks were close to zero. A one-sample t-test ($\mu = 0$) was performed on each subject's 15 scores ($p > .05$, one-tailed). One subject's scores did not differ significantly from zero, and they were excluded from further analysis. To balance the effect of this exclusion, the subject with the lowest mean ϕ_{adj} in each of the other three conditions was also excluded. These four subjects were not included in any subsequent analysis.

This particular exclusion criterion was chosen because it indicated for the remaining subjects that their set of 15 classifications were reliably more consistent with the category structure that we had defined than what would be expected by chance. This criterion does not demand that a subject is above chance on any particular block, just that his or her set of classifications are, as a whole, reliably above chance. However, it does allow us to consider only those subjects whose classifications are in accordance with the categorical structure we have imposed. Of course it is possible that such a criterion might exclude the vast majority of subjects if, for example, the way we defined our stimuli was totally at odds with how people perceived or grouped them. However, this does not appear to be the case for this experiment.

Figure 3a shows the mean ϕ_{adj} in all four conditions for each of the 15 blocks of the classification phase. A mixed-design analysis of variance (ANOVA), with one within-subject variable (block, 15 levels) and two between-subject variables (preexposed vs. non-preexposed and two categories vs. four categories) was performed, revealing three effects. First, a main effect of category indicated that classification was less consistent in the four-category conditions than in the two-category conditions, $F(1, 40) = 21$, $p < .001$. Second, a main effect of block revealed that consistency improved across blocks, $F(14, 560) = 11$, $p < .001$. Third, preexposed subjects were initially more consistent than non-preexposed subjects, an effect that diminished as the classification phase proceeded. This was indicated by a significant Preexposure $\times$ Block interaction, $F(14, 560) = 2.8$, $p < .05$, after a conservative correction for non-sphericity (Greenhouse & Geisser, 1959). No other effects approached significance in this analysis, $p > .13$ in all cases.

Following the significant Preexposure $\times$ Block interaction, a simple effects analysis revealed a significant difference between preexposed and non-preexposed subjects at Blocks 1, 2, 3, and 4, $F(1, 40) = 8.6$, 5.5, 4.9, and 5.0, respectively, $p < .05$ in all cases. A significant effect of block was found for non-preexposed subjects, $F(14, 560) = 12$, $p < .001$. The effect of block in the preexposed conditions was marginally significant, $F(14, 560) = 1.7$, $p = .06$.

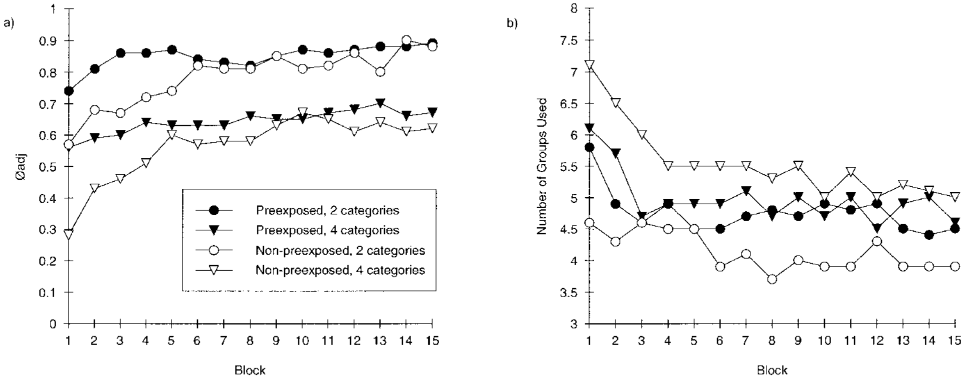

FIG 3. Mean consistency of subjects' free-classification judgements in the four conditions of Experiment 1a (as indexed by adjusted Cramér's ϕ), and the mean number of keys that they used.

The other main source of information available is the number of groups used by our subjects. Figure 3b shows the mean number for all conditions in each of the 15 blocks. A mixed-design ANOVA revealed two significant effects. First, a significant effect of block was found, indicating that subjects used fewer keys as the classification phase proceeded, $F(14, 560) = 5.1$, $p < .001$. Second, a main effect of category was seen, such that subjects in the four-category conditions used more keys than did subjects in the two-category conditions, $F(1, 40) = 4.6$, $p < .05$. In addition, there was some evidence of an interaction between category and preexposure, although it failed to reach significance, $F(1, 40) = 2.4$, $.1 < p < .15$. No other effects or interactions approached significance, $p > .12$ in all cases.

The marginal interaction and inspection of the means led us to the post hoc hypothesis that the difference between the number of keys used in two- and four-category problems was limited to the non-preexposed conditions. A simple effects analysis provided some tentative support for this notion—a significant difference was found between the number of groups used in the two-category–non-preexposed and four-category–non-preexposed conditions, $F(1, 40) = 6.8$, $p < .05$, whereas no other effects were significant, max $F(1, 40) = 1.4$, $p > .2$. This result is reported not because we are convinced of its reliability but because it is interesting and might be worthy of investigation in future experiments.

Discussion

The result of central importance is that subjects exposed to examples from the categories they would later be classifying produced, at least initially, more consistent classifications than did those who were not preexposed. We expected to see some benefit of preexposure early on in classification as this would be congruent with a perceptual learning effect (i.e. an increase in stimulus discriminability) analogous to that seen in McLaren et al. (1994). The effect would be transient because the classification experience itself would also allow perceptual learning, and so the effects of preexposure would be progressively diluted as classification progressed.

It would, however, be premature to accept this explanation because there are a number of potential alternative explanations that cannot be discounted at this point. For example, preexposed subjects have more experience at making decisions about black and white patterns under time pressure than non-preexposed subjects, which may lead to some sort of non-specific speed up. In other words, the beneficial effect of preexposure may be entirely due to increased familiarity with the general task, rather than increased familiarity with the stimuli. It is also possible, although perhaps unlikely, that subjects are covertly free classifying stimuli in the preexposure phase. Hence they would be better than non-preexposed subjects at classifying the stimuli on Block 1 of the free-classification phase because they are effectively on at least Block 6 at this point. We consider this second alternative to be unlikely given the highly demanding nature of the running-recognition task and the fact that classifying stimuli into similarity-based groups would, at best, make the task no easier. However, neither alternative can be rejected on the basis of evidence at this point—this will be one of the issues confronted by the two experiments to follow.

The results of the current experiment also give us some interesting information about the free-classification process. One of the immediately apparent effects was that subjects faced with a two-category structure were more consistent than those faced with a four-category structure. This is not surprising given that the four-category version of the task is intrinsically more difficult, requiring more information on the part of the subject to achieve perfect performance. However, subjects faced with a four-category structure also use more groups. This confirms that the form that subjects' classifications take is sensitive to fairly subtle differences in stimulus structure, and that subjects respond appropriately to these changes.

In all conditions, the number of groups used decreases as classification experience and classification consistency increase. This is clear evidence of the subjects learning something through classifying the stimuli. They are converging on a better degree of agreement with the programmed category structure in terms of both number and consistency of groupings. The non-significant interaction between preexposure (preexposed vs. non-preexposed) and categories presented (two or four categories) seen in the groups-used data, along with the associated trends in the means and simple effects, suggest that preexposure may modify this relationship in a non-trivial way. However, it would be unwise to overplay the importance of this result without further, more reliable, data.

EXPERIMENT 1B

The purpose of this experiment was to provide a sophisticated control for Experiment 1a. The basic idea was to replicate Experiment 1a in all aspects except the method of constructing examples from prototype patterns, this time using a method that we hypothesized would not lead to a significant beneficial effect of preexposure. If this hypothesis turned out to be correct, then we would have demonstrated that the original preexposure effect was dependent on specific properties of the stimulus rather than increased familiarization with the general task—hence supporting a perceptual learning explanation.

The choice of an alternative example-construction method was based on a series of informal conjectures that had previously found some support in a different task. Previous studies have shown that the effect of inverting a stimulus on a subject's ability to identify it as previously seen is dependent on both familiarity with the stimulus class (Diamond & Carey, 1986) and the type of stimulus used (Yin, 1969). In a recent paper, McLaren (1997) suggested that such results might be partially understood in terms of perceptual learning if one assumed perceptual learning was only effective for "prototype-based" stimuli. Prototype-based stimuli were defined as coming from a category of examples, which, if you averaged the set on a point-by-point basis, gave a pattern that was also an example of the category. Hence, size-matched pictures of faces would be an example of prototype-based stimuli but pictures of landscapes would not. McLaren went on to provide evidence that familiarity modulates the effect of inversion on the discriminability of pairs of chequerboard patterns, but only if the familiarized set was prototype-based.

The non-prototype-based stimuli used in McLaren (1997) were created from base patterns by randomly reordering the rows ("shuffled" stimuli) rather than giving each square an independent chance of reversing its colour. They are non-prototype–based in the sense that averaging these shuffled stimuli square by square would lead to a stimulus of 16 grey vertical bands of various intensities rather than another, prototypical, chequerboard pattern. If this change in the method used to construct examples were to affect significantly the impact of preexposure on free-classification decisions then we would have evidence that would be hard to explain on the basis of general familiarity with either the task requirements, or black and white stimuli as a general class.

Method

Subjects and Apparatus

The subjects were another 48 adults, again mostly graduate and undergraduate students from Cambridge University, aged between 18 and 30 and paid for their participation. The apparatus was that used in Experiment 1a, except that both computers were Risc PC600s connected to AKF60 monitors.

Stimuli

For each of the four experimental conditions, the 12 sets of category prototypes produced in Experiment 1a (1 set per subject, 12 subjects per condition, hence 12 sets per condition) were allocated in a random order to the subjects in this experiment. We use the term prototype here for convenience in designating stimuli that did act as prototypes in Experiment 1a, but define categories without being prototypes for them in this experiment. Hence, in terms of the prototypes, for each subject in Experiment 1a there was a corresponding defining stimulus in this experiment. Category exemplars were produced by randomly reordering the 16 rows of the appropriate defining stimulus. For example, the 9th row of the defining stimulus might appear as the 1st row of the exemplar, the 3rd as the 2nd, the 15th as the 3rd, and so on. This defines categories made up of sets of exemplars all of which have an equal standing as category members. There is no prototype, as the central tendency of the exemplars of a given category will be a columnar arrangement of greys rather than a chequerboard and, by definition, a prototype has to be a member of the set of examples for which it is a prototype.

Procedure

The procedure was identical to that in Experiment 1a.

Results

Experiment 1b

Separate one-sample t tests ($\mu = 0$), were run on subjects' adjusted Cramér's phi scores on the last block of each of the four conditions. In all conditions, scores were significantly greater than zero, $t(11) = 2.7$, 4.4, 3.5, and 3.2 for the two-category–non-preexposed, four-category–non-preexposed, two-category–preexposed and four-category–preexposed conditions, respectively, $p < .05$ in all cases. As before, inspection of individual subject data revealed some subjects whose scores across 15 blocks were close to zero. A one-sample t test ($\mu = 0$) was performed on each subject's 15 scores ($p > .05$, one-tailed). Subjects whose scores did not differ significantly from zero were excluded. As more subjects were removed on this basis in some conditions than in others, further subjects were excluded until the group sizes were equal. The subjects selected for exclusion were those with the lowest mean ϕ_{adj}. Eight subjects were removed in all (two per condition), and were not included in any subsequent analysis.

Figure 4a shows the mean ϕ_{adj} in all four conditions for each of the 15 blocks of the classification phase. A mixed-design ANOVA, with one within-subject variable (block, 15 levels) and two between-subject variables (preexposed vs. non-preexposed and 2 categories vs. 4 categories) revealed three effects. First, classification is less consistent in the four-category conditions than in the two-category conditions, $F(1, 36) = 7.9$, $p < .01$. Second, consistency improves across blocks, $F(14, 504) = 2.6$, $p < .01$. Third, consistency was lower in the preexposed conditions than in the non-preexposed conditions, $F(1, 36) = 4.8$, $p < .05$. The interactions were non-significant, $p > .3$ in all cases.

Figure 4b shows the mean number of groups used for all conditions in each of the 15 blocks. A mixed-design ANOVA revealed two effects. First, subjects used fewer groups as the classification phase proceeded, $F(14, 504) = 8.0$, $p < .05$. Second, preexposed subjects used fewer groups than did non-preexposed subjects, $F(1, 36) = 5.0$, $p < .05$. The trend for subjects in the four-category conditions to use fewer keys than subjects in the two-category conditions is non-significant, $F(1, 36) = 2.8$, $p = .1$. No other effects or interactions were significant, $p > .15$ in all cases.

Comparison of Experiments 1a and 1b

Given that the designs of Experiments 1a and 1b differ only in their stimulus structure, it is useful to compare them directly in one overall analysis. Comparison of Figures 3 and 4 suggests some important differences in the ϕ_{adj} data between Experiments 1a and 1b. A mixed-design ANOVA with one within-subject variable (blocks, 15 levels) and three between-subject variables (Exp. 1a vs. Exp. 1b, two vs. four categories, and preexposed vs. non-preexposed) revealed a number of significant results. Consistency was lower in Experiment 1b than in Experiment 1a, $F(1, 76) = 137$, $p < .001$, and consistency increased faster in Experiment 1a than in Experiment 1b, as indicated by a significant Experiment $\times$ Block interaction, $F(14, 1{,}064) = 2.1$, $p < .05$, after correction

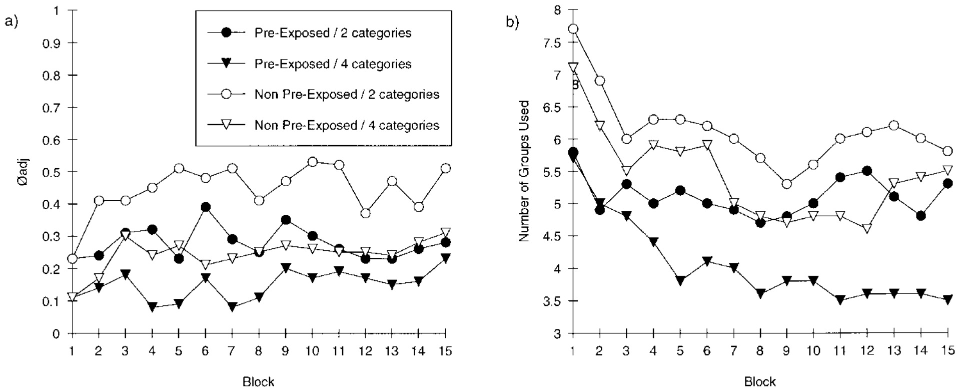

FIG. 4. Mean consistency of subjects' free-classification judgements in the four conditions of Experiment 1b (as indexed by adjusted Cramér's ϕ), and the mean number of keys that they used.

for non-sphericity. The difference between the effect of preexposure in the two experiments is significant, as indexed by an Experiment × Preexposure interaction, $F(1, 76) = 6.2$, $p < .05$. The increase in consistency across blocks and higher consistency for two-category conditions, seen in both experiments, remained significant when collapsing across them, demonstrated by a main effect of block, $F(14, 1{,}064) = 9.5$, $p < .001$, and a main effect of category $F(1, 76) = 26$, $p < .001$, respectively. There was also a significant Preexposure × Block interaction $F(14, 1064) = 2.6$, $p < .05$, after correction for non-sphericity. No other effects or interactions were significant, $p > .4$ in all cases.

From our account of the effects in this pair of experiments, the first block has particular significance because it gives the assessment of perceptual learning in preexposure, which is least contaminated by possible perceptual learning during free classification. Given the significant Preexposure × Block interaction in the above analysis, and given the importance we attach to the first block, we performed a second analysis of variance on the data from the first block only. Other than the main effects of experiment and category already found in our analyses thus far, there was a main effect of preexposure, $F(1, 76) = 4.6$, $p < .05$, reflecting the overall dominance of the beneficial effect of preexposure found in Experiment 1a, and also an Experiment × Preexposure interaction, $F(1, 76) = 5.0$, $p < .05$, indicating that the effects of preexposure on performance in the first block differed significantly in the two experiments. No other effects approached significance, $p > .2$ in all cases.

A final mixed-design ANOVA was performed, this time on all 15 blocks of the groups-used data from Experiments 1a and 1b. Three effects were revealed. First, the effect of the number of categories present on groups used was found to differ significantly between experiments—the Experiment × Category interaction gives an $F(1, 76) = 6.9$, $p < .05$. Second, the effect of preexposure on groups used seemed to be different in the two experiments, although the Experiment × Preexposure interaction was only marginally significant, $F(1, 76) = 3.6$, $p = .06$. Third, unsurprisingly, the decrease across blocks of the number of groups used seen in both experiments remained significant when collapsing across them, $F(14, 1064) = 12$, $p < .05$. No other effects were significant, $p \geqslant .08$.

Discussion

The results of Experiment 1b sugegst that it was more difficult to classify the stimuli used in that experiment than those used in Experiment 1a. Overall consistency was lower, learning was less rapid, and differences in the number of categories programmed was not reflected by differences in the number of groups that subjects used. The trend for subjects to use fewer groups in a four-category problem than in a two-category problem, although not reliable, is also slightly concerning. Nevertheless, we would argue that subjects were able to learn to classify these stimuli in a way that somewhat reflected the programmed category structure. The improvement in consistency over blocks was significant, and performance on the final block was significantly above chance in all conditions. We were not surprised that subjects found free classification with these stimuli more difficult, as the variation in stimuli drawn from a given category was necessarily much greater than that in Experiment 1a, and caution must be exercised in comparison of the two experiments for this reason. Even so, we do believe that Experiment 1b provides valuable information that helps us understand the results of Experiment 1a as well, particularly as far as preexposure is concerned.

The effects of preexposure in Experiments 1a and 1b are quite different as far as the consistency measure is concerned. Preexposure enhances consistency in Experiment 1a, particularly on the first few blocks, whereas it lowers consistency in Experiment 1b, particularly after the first few blocks. Such a result does not seem consistent with the notion that preexposure simply increases familiarity with the general task requirements or increases the ability of subjects to make rapid decisions about small black and white stimuli in general. If this were the case then we would expect preexposure to be beneficial in both Experiment 1a and Experiment 1b. Instead, the effects of preexposure appear to depend on the precise stimulus structure employed in the experiment.

It also seems difficult to explain the difference between Experiments 1a and 1b by assuming covert free classification of examples in the preexposure phase. Consistency improves with increasing classification experience in both Experiment 1a and Experiment 1b. Hence, if subjects were covertly free classifying, we would expect preexposure to be beneficial in both cases. We would not predict that preexposure should make the subjects' responses significantly less consistent with the programmed category structure. However, this is exactly what happened in Experiment 1b. The result is all the more remarkable given that overall consistency is not very good—if there had been any tendency for preexposure to improve performance it should have been evident in this experiment.

Preexposure leading to worse performance—the complement of perceptual learning—is an effect that has been seen before in humans and other mammals, often going under the title "latent inhibition''. However, it has previously been assumed that adult humans must direct their attention away from the presented stimuli to show this effect (see Lubow, 1989, for a review). Although subjects in this experiment are doing different tasks in the preexposure and free-classification phases, it seems clear that they must direct their attention towards the stimuli to perform (reliably) above chance during preexposure.

This point is worth emphasizing because, as far as we are aware, this experiment constitutes the first evidence of latent inhibition in adult humans without a concurrent distracting task. The implications of this will be considered in more detail later, but two points must be made immediately clear. First, in describing this effect as latent inhibition we are simply referring to the fact that a retardation in learning as a result of simple stimulus exposure has been demonstrated. We are not, at this point, making a statement about the process that is assumed to underly the result (this will come later). Second, although we did not expect prior to running the experiment that preexposure should significantly reduce the consistency of free-classification decisions, the result is, in fact, predicted by the application of a theory forwarded by one of the authors several years before the experiment was run (McLaren, Kaye, & Mackintosh, 1989). This will be discussed after reporting the final experiment in this investigation.

The effects of preexposure on key use seen in Experiment 1b is different to that seen in Experiment 1a. In Experiment 1b, preexposure reduced the number of keys used by subjects, irrespective of the number of categories programmed. Although the exact effect of preexposure on key use in Experiment 1a is somewhat unclear, a comparison of the two experiments suggested that the effects seen in the two experiments were different. Even without comparison to Experiment 1a, Experiment 1b provides clear evidence that preexposure can affect the classification system that a subject decides to adopt. Classifications using different numbers of groups are obviously not equivalent. A tentative explanation of the changes in the number of groups used as a result of preexposure is offered in the General Discussion.

EXPERIMENT 2

Taken together, Experiments 1a and 1b are only the second demonstration in humans that perceptual learning is contingent on both preexposure and stimulus structure (McLaren, 1997, being the first). In our next experiment we further extend this result by manipulating stimulus structure in a different way. Previous studies with rats have demonstrated that a beneficial effect of stimulus preexposure is dependent on the exposed stimuli being fairly similar. For example, in a taste-aversion procedure Mackintosh, Kaye, and Bennett (1991) found that simple exposure to two compound flavours with a common component (lemon–saline and lemon–sucrose) led to an enhanced ability to discriminate them later, compared to non-preexposed controls. However, preexposure to two simple flavours (saline and sucrose) had no significant effect on their discriminability.

Chamizo and Mackintosh (1989) found a similar result with a different procedure. They demonstrated that if rats had to discriminate between one arm of a maze that had a black rubber floor and another that had a red sandpaper floor, then simple prior exposure led to a slight facilitation. However, if the walls of one arm were then painted white, and the walls of the other were painted black (presumably reducing the similarity of the two arms) then preexposure led to a worsening in discrimination performance.

In our final experiment, we investigated whether an analogous effect could be demonstrated in the free classification of preexposed chequerboard patterns. We varied the similarity (assessed on a square-by-square basis) of the prototype chequerboard patterns used to create the examples presented to subjects. The prediction was that a larger perceptual learning effect would be seen for subjects in the condition with more similar prototypes. Note that if preexposure causes changes in stimulus discriminability then the size of its effect should be dissociable from overall performance on the free classification task. For example, free-classifying stimuli produced from very similar prototypes might be more difficult than free-classifying stimuli from moderately similar prototypes, but performance on the former task might be improved more by preexposure.

We decided that the most useful, least contaminated data on the effect of preexposure on the consistency of free-classification judgements would be found in the first block of free classification. There were two reasons for this conclusion. First, the results of Experiment 1a indicated that the beneficial effects of preexposure were only detectable early in free classification, and were numerically largest in the first block. Second, in the first block any effects of perceptual learning due to preexposure would be minimally contaminated by any effects of perceptual learning in free classification. Despite the importance we attach to Block 1, we decided there was little to be gained (apart from a few minutes of our subjects' time) by not running the remaining 14 blocks run in the previous experiments.

To recap, one purpose of the experiment described below was to test the influence of prototype similarity on preexposure effects. However, to increase efficiency in data collection we decided to include, in the same experiment, a further test that the effects of preexposure are stimulus specific. If preexposure effects are stimulus specific, one might expect that preexposure to entirely random chequerboard patterns would not be as beneficial as preexposure to patterns that were all validly produced from the prototype patterns. At the absolute limit, we might expect no effect of preexposure to random patterns because their construction bears no statistical relationship to the construction of the patterns that are to be classified. Hence one might test whether the nature of stimulus construction was the only important factor in these experiments by comparing preexposure to random chequerboard patterns with no preexposure at all.

Method

Subjects and Apparatus

The subjects were a further 72 graduate and undergraduate students from Cambridge University, aged between 18 and 30 and paid for their participation. The apparatus was the same as that used in Experiment 1b.

Stimuli

All stimuli were again 16 × 16 chequerboards, but there were a number of varieties—overlap, complement, and random. *Overlap* stimuli were constructed in the same way as the two-category stimuli in Experiment 1a, in other words, by the addition of random noise to two prototype base patterns that shared 50% of squares. The actual prototypes were different to those used in Experiments 1a and 1b, but the method of construction was the same. *Complement* stimuli were created by the addition of random noise to two base patterns that were the exact negative image of each other (see Figure 5). One base pattern consisted of 128 black squares and 128 white squares randomly placed, the other was created by reversing the colour of each square in turn. For both overlap and complement stimuli, the nature of the random noise was the same as that in Experiment 1a; each square had an independent 5% chance of having its colour reversed. It seems reasonable to assume that the "negative image" relationship between the complement stimuli is not obvious to subjects (or, at least, far less obvious than it appears in Figure 5). This is because the level of noise added means that the subjects rarely, if ever, see the base patterns, and because patterns are presented sequentially rather than simultaneously.

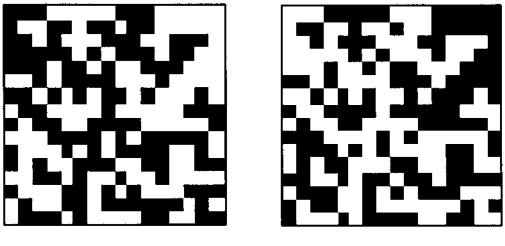

FIG. 5 Two examples of the type of chequerboard stimuli used in Experiment 2. The two prototype patterns, which are the exact inverse of each other, are shown. The rectangle outlines enclosing the stimuli represent the beginning of the mid-grey background against which they were presented; they are not part of the stimulus itself and were not presented to subjects.

Each of the *random* stimuli was simply a 16 × 16 chequerboard formed by the random arrangement of 128 black squares and 128 white squares.

Procedure

The experiment was of a 3 × 2 factorial between-subjects design, the factors being preexposure (non-preexposed, random preexposed or related preexposed) and stimulus type (overlap or complement). The overlap-preexposed and overlap–non-preexposed conditions were a replication of the two-category conditions of Experiment 1a—they consisted of 15 blocks of two-category free classification, preceded either by 5 blocks of running recognition (preexposed) or an unrelated experiment of approximately the same duration (non-preexposed). The complement-preexposed and complement–non-preexposed conditions were identical to the corresponding overlap conditions except that the complement stimulus type was used in both the running-recognition and free-classification phases. In the random-preexposed conditions each of the twice-presented stimuli in the running-recognition phase was of the random type; the stimuli in the free-classification phase were either complement or overlap, depending on condition. Conditions in which preexposure stimuli and free-classification stimuli are created in the same way are referred to as *related-preexposure* conditions, to draw the distinction between them and the *random-preexposure* conditions.

Results

Although we were predominantly interested in performance on the first block in this experiment, for the sake of consistency and completeness we applied the same tests of final performance and the same exclusion criteria as those used in previous experiments. Separate one-sample *t* tests ($\mu = 0$) were run on subjects' adjusted Cramér's phi scores on the last block of each of the six conditions. In all conditions, scores were significantly greater than zero, $t(11) = 17$, 11, and 13, for the non-preexposed, random-preexposed and related-preexposed overlap-stimulus conditions, and $t(11) = 5.8$, 11, and 7.4 for the corresponding complement-stimulus conditions, $p < .05$ in all cases. A one-sample *t* test ($\mu = 0$) was performed on each subject's 15 scores, $p > .05$, one-tailed. Subjects whose scores did not differ significantly from zero were excluded. As some conditions had more subjects removed on this basis than others, further subjects were excluded until the group sizes were equal. The subjects selected for exclusion were those with the lowest mean ϕ_{adj}. Twelve subjects were removed in all (two per condition—the same overall proportion as that in Experiment 1b) and were not included in any subsequent analysis.

The analysis of the data from this experiment proceeded in two stages. First, ANOVAs were carried out on the full data set to provide an overall picture. Following these, a number of specific analyses were performed to answer the questions that the experiment was designed to ask. Central to this second phase is the assumption that a significant main effect of group or a significant interaction in an overall analysis is not required to validate the results of a particular, planned, linear contrast. All that is required is that the specific hypothesis that one wishes to test could have been and was formed prior to collecting and inspecting the data. In the past, psychologists and statisticians have sometimes queried this assumption, but current thinking is that it is valid. Howell (1992, e.g. p. 338) or Wilcox (1987) may be consulted for an extended discussion of this point. The specific tests that we report relate directly to specific predictions made prior to running the experiment and made on the basis of prior experimental evidence.

Overall Analysis

Figure 6 shows the mean ϕ_{adj} scores for each of the 15 blocks in each of the six conditions in this experiment. Figure 7 presents the number of groups that subjects used in the same format.

A mixed-design ANOVA with one within-subject factor (block, 15 levels) and two between-subject factors (stimulus type, 2 levels [overlap or complement] and preexposure type, 3 levels [non-preexposed, random-preexposed and related-preexposed]) was run on the subjects' consistency scores. Two effects were found. First, consistency increased as

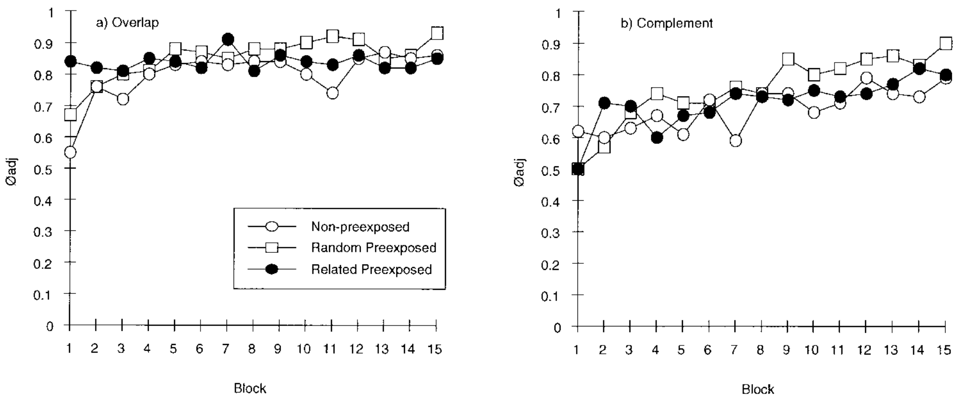

FIG. 6. Mean consistency of subjects' free-classification judgements (as indexed by adjusted Cramér's ϕ) in the non-preexposed, random-preexposed, and related-preexposed conditions of Experiment 2. The left panel shows the results for the overlap-stimulus type, the right panel results for the complement-stimulus type.

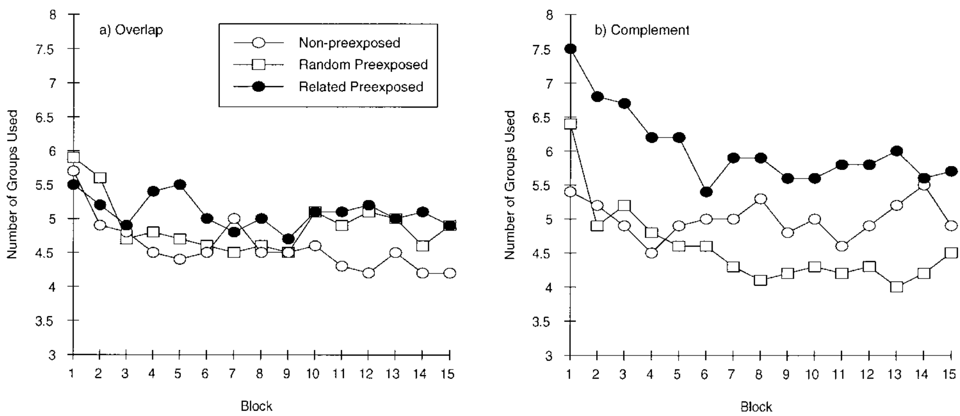

FIG. 7. Mean number of groups used by subjects in the non-preexposed, random-preexposed, and related-preexposed conditions of Experiment 2. The left panel shows the results for the overlap-stimulus type, the right panel results for the complement-stimulus type.

classification proceeded, as indicated by a main effect of block, $F(14, 756) = 10$, $p < .05$. Second, consistency scores were reliably lower in the complement condition than in the overlap condition, $F(1, 54) = 4.4$, $p < .05$. This latter effect is perhaps surprising and will be considered in detal in a later section. No other main effects or interactions approached significance, $p > .15$ in all cases.

An equivalent ANOVA was peformed on the number of groups subjects used, and two effects were found. First, the number of groups subjects used fell as classification proceeded, $F(14, 756) = 8.1$, $p < .05$. Second, there was a reliable effect of preexposure on the number of groups used, $f(2, 54)=3.5$, $p<.0.5$. However, a post hoc Tukey test failed to find a reliable difference between any of the three pairs of conditions, $p > .05$ in all cases. It is therefore hard to characterize clearly the locus of this unpredicted and unexpected effect. No other effects in the overall analysis were reliable, $p > .1$ in all cases.

Specific Tests

The Effect of Random Preexposure. There are two related questions. First, does random preexposure produce reliably different consistency scores to no preexposure? The answer given by a planned contrast of the two appropriate group means is no, $F(1, 54) = 0.82$, $p > .3$. Second, do random preexposure and no preexposure produce reliably different number of groups in subjects' classifications? The answer, again given by a planned contrast of the two group means, is no, $F(1, 54) = 0.02$, $p > .5$.

The Effect of Preexposure on Consistency Scores in Block 1 of Free Classification. There were three specific questions. First, do consistency scores in Block 1 of the related-preexposed–overlap condition differ reliably from consistency scores in Block 1 of the non-preexposed–overlap condition? The answer, given by a planned contrast between the two appropriate Block 1 group means, is yes, $F(1, 54) = 4.1$, $p < .05$. Second, do consistency scores in Block 1 of the related-preexposed and non-preexposed complement-stimulus conditions reliably differ? The answer is no, $F(1, 54) = 0.64$, $p > .4$. Third, is there a reliable difference between these two differences? This can be answered by an appropriate planned linear contrast between the four appropriate Block 1 group means, and the answer is yes, $F(1, 54) = 4.0$, $p = .05$.

Discussion

Preexposure was shown to increase the consistency of subjects' judgements when the two underlying prototypes were, on a square-by-square basis, 50% similar but not when they were 0% similar on the same basis. The conclusion we wish to draw is that this primitive similarity measure is a useful one and that the effect of preexposure on consistency is modulated by the similarity of the underlying prototypes on this measure. Specifically, a beneficial effect of preexposure is seen only if the prototypes are fairly similar.

The validity of this conclusion clearly depends on the validity of the similarity measure. It would seem, intuitively, that the perceived similarity of two visual patterns is determined by more than the point-by-point agreement of their intensities. The presence of common edges and a number of other relational and configural properties are probably also important and, indeed, we believe this to be the case. However, for our purposes it is not crucial that the number of squares shared should provide a complete description of how similar two chequerboard patterns are—it must simply provide a good first approximation to this. Although it seems unlikely that *squares shared* is a perfect index of the similarity of two chequerboards, it does seem likely that, as a general rule, the fewer squares two patterns share the less similar they will appear.

Some evidence for the latter point is provided in an experiment by McLaren, Bennett, Guttman-Nahir, Kim, and Mackintosh (1995, experiment 3). They investigated the perceived similarity of a varied set of 16×16 chequerboard patterns created by the addition of random noise to two prototypes that shared approximately 56% of squares. Similarity was indexed by the mistakes made in an identification-learning task. A one-dimensional scaling solution of the similarity matrix accounted for 90% of the variance, and the rank ordering of stimuli on that dimension correlated very highly, $r_s = .94$, with the rank ordering of stimuli on the basis of the number of squares they shared. In other words, squares shared appeared to be a reasonable estimate of similarity. Further details of this standard method of assessing psychological similarity space may be found in Nosofsky (1986) or Shepard (1987).

Overall, it seems that our initial interpretation was basically valid. To reiterate, the effect of preexposure on the consistency of free-classification judgements is modulated by the similarity of the underlying prototypes, and a beneficial effect is seen only if the prototypes are sufficiently similar. This provides further evidence that the effect of preexposure is highly dependent on the exact nature of the stimulus construction.

Additional converging evidence is provided by the other major result of this experiment. In terms of its effect on consistency, preexposure to chequerboard patterns that are a random arrangement of an equal number of black and white squares is not reliably different to the absence of preexposure. This is an important result because in many ways these random chequerboards are very like the patterns that subjects will later classify. Indeed, any random pattern can be thought of as a member of some chequerboard category. The difference between related and random preexposure is that in related preexposure all chequerboards are created from a small set of base patterns, which are the same as the set used to create the chequerboards used in free classification. Hence, our result suggests that it is the relationship between the set of patterns presented in preexposure and the set of patterns presented in free classification that is of primary importance in determining the effect of preexposure. If other factors, such as general familiarity with chequerboard patterns or with the task requirements, were important in determining the effect of preexposure, then one would expect that the effect of preexposure to random chequerboard patterns would be different to the effect of no preexposure at all. This experiment fails to provide any reliable evidence for this position.

There are two further novel results in this experiment. First, the type of preexposure received—related, random, or none—reliably affects the number of groups that subjects use. The available data do not allow us to decide which types of preexposure are reliably different (none of the post hoc pairwise comparisons was significant). Nevertheless, the result provides further evidence that preexposure can affect the classification system that a subject decides to adopt. Second, subjects in the overlap conditions produce more consistent classifications than do subjects in the complement conditions. An explanation of this result will be given towards the end of the final section.

GENERAL DISCUSSION

This paper contributes three novel empirical findings concerning the effects of stimulus preexposure on human subjects. First, Experiment 1 establishes that the effects of preexposure on category learning depend on the stimulus structure of the categories in question; McLaren's (1997) analogous result was for within-category discriminations rather than the between-category discriminations considered here. Second, Experiment 1b is, as far as we are aware, the first demonstration of a retardation in learning as a consequence of simple preexposure in adult human subjects. Other demonstrations of this type of effect have relied on incidental or masked preexposure to generate retarded learning with adult humans. Third, Experiment 2 provides further support for the contention that the effects of preexposure are contingent on stimulus structure, and it is the first experimental demonstration in humans that stimulus similarity is a crucial factor in determining the consequences of stimulus preexposure.

Taken together, these findings rule out one class of explanation of the type of exposure-learning phenomena reported here. One might attempt to argue that the effects of exposure learning were due to subjects automatically paying more or less attention to familiar stimuli. Such an explanation would be in the spirit of conditioned attention theory (Lubow, 1989) or of certain accounts of negative-priming phenomena (Tipper, 1985). The fact that the precise nature of the stimulus structure of otherwise similar patterns can, by itself, determine whether familiarization leads to speeded or retarded learning makes such an account hard to sustain. The crux of the problem is that attention is assumed to be directed towards or away from stimuli as unitary objects. In the preexposed conditions of all our experiments, subjects have been familiarized, under the same running-recognition procedure in each case, with objects that are in some way similar to those that they later have to classify. However, no uniform beneficial or detrimental effect is seen. It seems difficult to explain this result in terms of shifting attention towards or away from stimuli as unitary wholes.

The results can be explained, however, by the extension of a model originally designed to explain perceptual learning and latent inhibition in rats and pigeons proposed by McLaren et al. (1989). It is important to emphasize that the theoretical discussion that follows is not a post hoc speculative account. The predictions of the McLaren et al.—hereafter MKM—model derive from a simple application of it to the new data in this paper, and the assumptions underlying this application are the same as those previously used to explain other human and animal data.

For a precise mathematical statement of the MKM model, the reader is invited to consult the original paper (McLaren et al. 1989). However, all that is required to understand the predictions of the model for the current experiments is an understanding of the basic underlying principles. Of central importance is the principle that all stimuli, however simple, are represented by a number of features or attributes—hereafter *elements*. As a result of repeated presentation and reliable co-occurrence, associations may form between these elements. The formation of associations is assumed to be determined by a simple error-correcting rule—the delta rule (McClelland & Rumelhart, 1985). This algorithm is similar in spirit to that used in the Rescorla-Wagner theory of Pavlovian conditioning (Rescorla & Wagner, 1972).

Overlying this associative learning system is a mechanism (the modulator) that generates changes in the salience (associability) of elements. As the network of associations becomes able to predict the occurrence and non-occurrence of an element (i.e. as the error term in the learning rule reduces) the modulator reduces the salience of that element and hence the element is less able to enter into new associations.

For our purposes, this is the key feature of the model. In situations where elements reliably predict each other, repeated exposure to those elements leads to a reduction in their salience and hence a reduction in the rate at which new associations can form. Crucially, preexposure often leads to differential changes in salience of different elements. The extent to which the salience of an element changes is, at limit, entirely determined by the extent to which it is predicted by other elements. However, the overall probability with which an element occurs also generally affects the salience of its representations—high-probability elements occur more frequently, and hence associations to their representations form more quickly than do associations to the representations of low-probability elements. Previous discussions of the application of the model (e.g. Mackintosh et al., 1991; McLaren, 1997; McLaren et al., 1989, 1994) have concentrated on the relative frequency of occurrence of elements, but this does not alter the fact that the model's predictions are determined by the overall frequency of an element's occurrence as well as by the contingency between it and other elements.

In applying the MKM model to the current data, we assume that the prototypes of our chequerboard patterns are composed of a number of elements. Some of these elements are unique to a particular prototype, whereas some are shared with one or more of the other prototypes. When examples of catgories are produced from prototypes by the addition of random noise, as they are in Experiments 1a and 2, we assume that these examples are composed mainly of elements that make up the prototype but that the examples also contain some non-prototypical or "noise" elements. These assumptions are exactly the same as those made in the explanation of the results in McLaren (1997) and McLaren et al. (1994) and follow logically from the initial application of the model to single stimuli and pairs of stimuli (e.g. Mackintosh et al., 1991; McLaren et al., 1989). The question of how to apply the model to the form of stimulus construction used in Experiment 1b will be considered later.

An important aspect of these assumptions about stimulus representation in Experiments 1a and 2 is that the model is able to learn something about the categorical nature of the stimuli without actually categorizing them (either with or without the help of externally provided labels). The presence of single elements can be predictive of the experimenter-defined category of the stimulus that contains them, in the sense that the two pieces of information are positively correlated. The model cannot learn the predictive relationship directly when no category label is provided, but it can learn which subsets of elements reliably co-occur, which do not, and which predict the absence of others. Because category examples are composed of a set of reliably co-occurring elements, a single element that is diagnostic of category membership is more predictable by other elements than one whose occurrence is totally independent of the defined category structure (assuming that overall probability of occurrence is held constant). Taking this one step further, if an element is negatively correlated with a particular category then if for some reason it does occur in that category it will be in circumstances where it is specifically predicted not to be present. In other words, the model is sensitive to the correlation of elements with category membership even though category-membership information is never specifically provided. Again these properties of the model arise from a standard application of it to the current experiments.

To summarize, the salience of the representation of an element changes with exposure. The salience of a representation is the rate at which associations from it to other representations are learned. Salience is affected by both the frequency of occurrence of an element in conjunction with other elements and the extent to which occurrences are predicted by other elements. Higher frequency of occurrence leads to lower salience, as does higher predictability. Due to the categorical nature of the stimuli used in these experiments, for any given frequency of occurrence elements correlated with category membership are more predictable than those uncorrelated with category membership. In turn, elements negatively correlated with category membership are less predicted to occur in that category than uncorrelated elements.

Figure 8a illustrates schematically the effect of these co-determinants of salience in a two-category situation in our experiments. A small number of discrete points are plotted to keep the figure simple but, of course, probability of occurrence and correlation with category membership could potentially take any value within the ranges shown. Note that not all lines are of equal length; an element with no correlation to the presented category may have any frequency of occurrence, but an element perfectly correlated with category membership must, in the two-category conditions of our experiments, occur 50% of the time. Figure 8b shows how the *common and unique elements* Venn diagram used in the original presentation of the MKM model (McLaren et al., 1989) may be related to Figure 8a in a simplified situation where all elements are sampled. This Venn diagram illustrates a simple case where two stimuli (the two circles of the Venn diagram) are preexposed. The elements that the two stimuli share are *common* elements; they have a probability of occurrence of one and a correlation of zero with the occurrence of a particular stimulus. Elements exclusive to one stimulus are *unique* elements. In a two-stimulus situation, the occurrence of unique elements is perfectly correlated with the occurrence of the stimulus of which they are a part, and their overall probability of occurrence is .5 if the two stimuli are presented equally often. In the more complex situations of the current experiments, correlation with category membership and overall probability of occurrence can take values other than 0, .5, and 1. Figure 8a shows the full range for the two-category conditions.

We are now in a position to explain the effects of preexposure on the consistency judgements in Experiments 1a and 2. The assumption made is that free classification proceeds by the formation of associations from elements in the stimulus representation to category-level representations and that these associations form faster to more salient elements. The exact process by which this occurs must, at least in this current paper, remain unspecified. A fully adequate connectionist model of free classification is a worthy and substantial project in its own right, but not one that is undertaken here. However, if one accepts our basic conjecture about free classification, what predictions can be made?

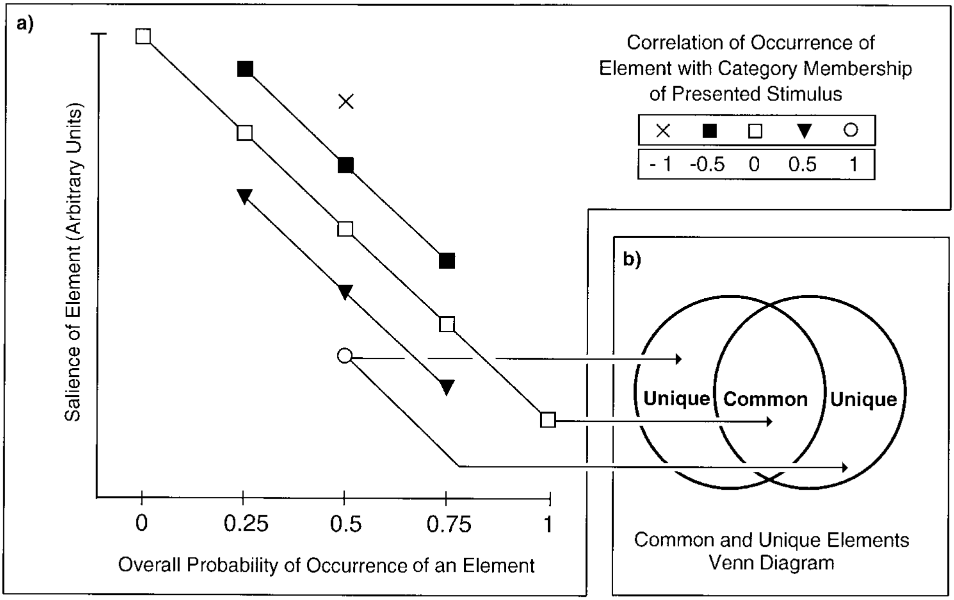

FIG 8. (a) The salience of representational elements in the McLaren et al. model (1989) as a function of the element's overall probability of occurrence, and the correlation of its occurrence with the category membership of the presented stimulus. (b) The relationship between certain points of the graph in Figure 8a and the common and unique elements Venn diagram used in the original presentation of the model (McLaren et al., 1989).

The following discussion will concentrate on the two-category case for the sake of both brevity and clarity.

In Experiment 1a, two types of element predominate. Ignoring noise for the moment, these are the *unique* elements, which are perfectly correlated with category membership and occur 50% of the time, and the *common* elements, which occur 100% of the time but are totally uncorrelated with category membership. As previously discussed, and as summarized in Figure 8, the unique elements are higher in salience than the common elements as a result of preexposure.

This may be made clearer by the following simplified two-stimulus example. Imagine a stimulus (stimulus A) with four elements. Two of these elements, *a* and *b*, are unique to stimulus A. The other two elements, *c* and *d*, are common to both stimulus A and a second stimulus (stimulus B). When stimulus A is first presented, all four elements occur together and are equally associated with each other. However, the common elements *c* and *d* occur together again when stimulus B is presented, so that their association is strengthened further. Over a series of preexposure trials, the common elements become more strongly associated than the unique elements, leading to lower salience for these common elements.

Thus we assume that, after preexposure, associations form preferentially to the unique elements. In the absence of preexposure, the salience of common and unique elements are initially the same. As the unique elements are the ones that determine the experimenter-defined category membership and are therefore the ones to which we want subjects to form associations, the consequence of preexposure is higher consistency of free-classification judgements.

If we now allow for noise, it has little impact on the salience of unique and common elements (as there is only 5% noise in any case). It can be seen from Figure 8a that noise elements are high in salience because their overall probability of occurrence is low and their correlation with category membership is either zero (noise added to common elements) or close to -1 (noise added to unique elements). The overall impact of noise elements is small, however, as they are very much in the minority. Hence the model predicts that preexposure is beneficial in this situation. The explanation of the effect of preexposure on consistency in Experiment 1a is essentially equivalent to the "differential latent inhibition of common and unique elements" account of perceptual learning in rats presented in McLaren et al. (1989).

The model also accounts for the result in Experiment 2 that the effect of preexposure on consistency is modulated by the similarity of the underlying prototypes. If the two prototypes are less similar, the number of common elements in the examples produced from them are fewer. As the difference in salience between common and unique elements is assumed to underly the basic preexposure effect of Experiment 1a, reducing the number of common elements should reduce the beneficial effect of preexposure. In the extreme case, if two stimuli share no common elements then preexposure simply makes all elements lower in salience, resulting in a detrimental effect of preexposure. To explain the similarity-dependence results of Experiment 2, we must simply assume that our manipulation was effective in reducing the proportion of common elements but did not eliminate them entirely. Both Mackintosh et al. (1991) and Chamizo and Mackintosh (1989) explain their analogous results by a similar application of the MKM model. Although in our experiment the pair of base patterns used to construct examples in the complement conditions were entirely dissimilar on a square-by-square basis, it seems unlikely that the similarity of two chequerboard patterns is determined solely on this basis. Hence it may be reasonable to assume that the base patterns still share some elements in common.

There are other results in Experiment 1a and 2 to be explained, but first we need an explanation of the most striking result in this paper—the fact that the preexposure has a detrimental effect on consistency in Experiment 1b. The MKM model provides an explanation which follows naturally from two assumptions about stimulus representation that we have already made. First, the number of matching squares in corresponding positions is a good first approximation to the similarity of two chequerboard patterns. Second, the similarity of two patterns increases with the number of elements that their representations share in common. If these arguments are valid then it seems likely that a significant proportion of the set of elements representing a chequerboard pattern are those that signal the colour of a square in a specific position on that chequerboard. We will assume that these *single square* elements respond to the presence of a particular shade—that is, they are sensitive either to black or to white but not to both.

It is the changes in salience that these single-square elements undergo that are taken to drive the detrimental effect of preexposure seen in Experiment 1b. We have already demonstrated that in the MKM model the relative salience of elements as a result of preexposure is that shown in Figure 8. Hence, to extract the predictions of the model for Experiment 1b, one needs to characterize the single-square elements in terms of their correlation with category membership and their overall probability of occurrence. This can be done by the application of the assumptions about stimulus representation that we have already made, as is demonstrated below.

Overall probability of occurrence is considered first. In Experiment 1b, all examples were constructed by the random rearrangement of the rows of a base pattern. Therefore, the probability with which a single-square element occurs is determined by the sum of the total number of black squares and the total number of white squares in the appropriate columns of the base patterns. A specific, two-category example may make this clearer.

Consider an element that represents the presence of black in, for example, the third square from the top in the fourth column from the left. The overall probability of this element being activated when a chequerboard pattern is presented is the number of black squares in column 4 of base pattern one, plus the number of black squares in column 4 of base pattern two, all divided by the total number of squares—that is, 32 (2×16, 16 being the number of rows in our chequerboard patterns).

The master pattern from which the base patterns are constructed is made up of an equal number of black and white squares. Hence, the most likely probability of occurrence for any single-square element is .5. Of course, other probabilities of occurrence are possible, although less likely. The likelihood of all probabilities of occurrence can be determined exactly from a binomial distribution, which for large N approaches a normal distribution. Therefore the overall probability of occurrence for single-square elements can be characterized by as an approximately normal (i.e. Gaussian) distribution centered on .5.

The correlation between the occurrence of a single-square element and the occurrence of an example from a particular experimenter-defined category is also determined by the number of black and white squares in columns of the base patterns. However, it is the difference between the number of squares of a particular colour in the two base patterns (rather than the sum) that determines correlation. Again, a couple of examples may make this clearer. For simplicity, the overall probability of occurrence is held constant at .5, and only elements representing the presence of black in particular squares are considered.

First, consider two base patterns (base pattern one and base pattern two). Both have 8 black squares in the fifth column from the right. An element representing the presence of black in, say, row 10 of that column will occur on half the presentations of base pattern one. However, it will also occur on half the presentations of base pattern two. Hence the element's correlation with the experimenter-defined category structure is zero.

Now consider a second situation. Base pattern one has 16 black squares in the fifth column. Base pattern two has no black squares in this column. Overall, the element will occur 50% of the time—the same as in the previous example. However, the element will only occur if an example from base pattern one is presented. It will never occur if an example from base pattern two is presented. Hence the occurrence of the element is perfectly correlated with category membership.

The master pattern from which the base patterns are composed has an equal number of black and white squares. Therefore, for any given column, the occurrence of 8 black squares in base pattern one and 8 black squares in base pattern two is the single most likely outcome. The occurrence of sixteen black squares in one pattern and zero in the other is extremely unlikely. As before, the exact probabilities can be derived from binomial distributions. Binomial distributions approximate to a normal distribution for large N. Hence, the correlation of the occurrence of single-square elements with the experimenter-defined category can be characterized by a probability distribution centred on zero and approximately Gaussian in shape.

To summarize, single-square elements in Experiment 1b can be characterized as follows—their overall probability of occurrence has a distribution with a mean of .5 and their correlation with category membership has a distribution with a mean of zero. As in Experiment 1a, the mean salience of elements is predicted to be lower after preexposure. The model predicts increased consistency as a result of preexposure in Experiment 1a, even though mean salience is lower, because salience changes differentially for diagnostic and non-diagnostic elements. Due to the nature of stimulus construction in Experiment 1a, the non-diagnostic (common) elements turn out to have a lower salience than the diagnostic (unique) elements. This leads to increases consistency because the diagnostic elements are preferentially associated to category representations.

The situation is very different in Experiment 1b. Whereas in Experiment 1a, relative frequency was the dominant factor in determining element salience, in Experiment 1b this is no longer the case. In Experiment 1b, it is correlation with category membership that plays the dominant role, because stimulus construction is such that element frequency is no longer confounded with category membership. In Experiment 1a, diagnostic elements occurred more frequently than non-diagnostic ones. However, in Experiment 1b if an element has a high correlation with a category its probability of occurrence is, on average, .5. If it has zero correlation, the mean probability of occurrence is still .5. If it is negatively correlated, the mean probability of occurrence is still .5. Correlation is not confounded with frequency in this experiment.

As noted earlier, elements that are correlated with category membership co-occur more often with other elements in that category than elements that are not positively correlated; this increased co-occurrence leads to stronger interassociations and hence lower salience. Hence, for any overall probability of occurrence, elements positively correlated with category membership are lower in salience than elements with zero correlation. In turn, the salience of elements with zero correlation is lower than the salience of those elements that have occurred in a category of which they are not generally predictive (i.e. negative correlation). This means that, as a result of preexposure, associations tend to form to the wrong elements first, the uninformative ones second, and the right ones third and last. In the absence of preexposure, associations form at the same rate to all three. Therefore, preexposure reduces the consistency of subjects' free classifications in this situation. It is worth emphasizing that, although the discussion leading up to this conclusion has been lengthy, it did not require any more assumptions than had already been made and supported.

The lack of any significant effect of random preexposure in Experiment 2 is also simply explained by the MKM model. Under random preexposure, the occurrence of any single-square element is independent of any other. Hence, these elements are unable to predict each other. (Although associations are formed on those occasions when elements occur together by chance, with repeated random presentations the MKM model predicts that these associations will asymptote at zero.) The result is no change in salience, so that if the behaviour of single-square elements is driving the effects of preexposure, then one would expect no significant effect under random preexposure.

The fact that subjects in the overlap conditions of Experiment 2 are more consistent overall than subjects in the complement conditions is more puzzling. We have argued that the stimuli in the overlap conditions have more elements in common than those in the complement conditions. Most learning theories would predict the opposite result—that the less similar two stimuli are the easier they are to discriminate. We are unsure how to explain this result at present. As we have argued before, our measure of similarity (squares shared) is almost certainly incomplete. Clearly, there is some relationship between a chequerboard and its negative image, which goes beyond the fact that they share no elements in common on a square-by-square basis. Equally, this relationship is a difficult one to capture in representational terms—it seems different, more abstract, than simple feature overlap. It may be that this relationship is, in some way, responsible for overall poorer consistency in the complement conditions. However, as we have no further evidence to bring to bear at this point, further speculation seems idle.

Despite such unanswered questions, the extent to which a theory developed to explain latent inhibition and perceptual learning in animals transfers to the human domain is still one of the most remarkable aspects of our results (with the proviso that the results for human subjects are consistent with purely differential salience change). In particular, the ability of the theory to explain when faster and slower learning should occur as a result of preexposure is impressive. It suggests that there may be more scope for integrating theories from human and comparative psychology.

The above discussion has been limited to the consistency of free-classification judgements, but the results of the experiments reported are not limited in this way. They also demonstrate that preexposure can change the type of classifications that subjects use in stimulus-specific ways. In Experiment 1a, the results were not clear cut, but inspection of the means suggests that preexposure reduced the number of groups subjects used in a four-category problem but increased it in a two-category problem. In Experiment 1b, preexposure decreased groups used, irrespective of problem type, and there was some evidence this was a reliably different pattern of results to that seen in Experiment 1a. In Experiment 2, a significant effect of preexposure on groups used was found, but again its exact nature was difficult to characterize clearly. Intriguingly, inspection of the means in this case suggests that the biggest effect was in the preexposure of complement stimuli, where no reliable effect on consistency was seen. It may come as no surprise that we believe that some of the effects of preexposure on the number of groups used are explicable by changes in elemental salience. However, we had no clear predictions concerning specific changes in groups used prior to running these experiments and have, as yet, performed no empirical tests of the hypotheses we now have. It therefore seems inappropriate to expand on them here.

Although such theoretical and empirical development is beyond the scope of the current paper, it certainly needs to be made. In the current experiments we have shown that the phenomena of perceptual learning and category formation are intimately and intricately related. However, many otherwise successful models of categorization do not explain the type of category formation seen in our experiments. Examples include the generalized context model (Nosofsky, 1986), RULEX (Nosofsky, Palmeri, & McKinley, 1994), the array model (Estes, 1994), back-propagation networks (Rumelhart, Hinton, & Williams, 1986) and ALCOVE (Kruschke, 1992). These and many other models assume that categorization is based on item-specific feedback. Such feedback is not available in our experiments, yet subjects form categories consistent with the underlying, experimenter-defined structure. Neither is the theoretical account we present immune from such criticism. Although we have allowed that associations can form from stimulus-level representations to category-level representations without item-specific feedback, we have not provided an account of how this might happen (other than that the associations will form more rapidly to elements with higher salience).

However, the class of models discussed above would also have difficulty in capturing the representational development that simple exposure seems to produce. The generalized context model bases its predictions about categorization on the confusability of stimulus pairs derived from, for example, an identification experiment. The model could be applied to the effects of preexposure on categorization by running two identification experiments, one with an initial preexposure phase and one without. However, this would not provide a process explanation of perceptual learning, which is the goal of our investigations. ALCOVE has a mechanism by which attention to specific dimensions is modified, but this mechanism is dependent upon the network being presented with a category label. Additional processes would have to be specified if this model is to predict that simple exposure can affect later explicit categorization. Estes' array model has a parameter, or sometimes parameters, which express the confusability of stimulus features. One could reasonably postulate that preexposure changes their value. However this is, once again, some way from specifying a process by which it might occur. The remaining models mentioned so far appear to have no facility to deal with the effects of simple exposure.

In other words, these models provide a theoretical account of categorization but do not address themselves to category formation or exposure learning. Our model explains the exposure learning seen in our experiments, but does not specify a specific process for category formation in the absence of feedback. There are theories that specifically address category formation, but some (e.g. Ahn & Medin, 1992; Michalski & Stepp, 1983) cannot be applied to our data because they work on the premise that the entire to-be-classified stimulus set is fully accessible at the point of category formation. This is clearly not the case with our sequential classification procedures.

Two models that can be applied to our data are Anderson's rational model (Anderson, 1990, 1991) and COBWEB (Fisher, 1987). Both of these models work on the assumption that we form categories in such a way as to maximize the extent to which they predict properties of the individual stimuli. However, neither model includes any method by which stimulus representations can change as a result of simple exposure, a process that we have argued is central to our understanding of the results of the current experiments. The applications of standard, unsupervized connectionist models (e.g. Rumelhart & Zipser, 1986) to the current data suffers from a related problem. For such models, the preexposure and free-classification phases are not fundamentally different; preexposure simply provides more examples to classify. Hence preexposure should generally be beneficial when the free-classified and preexposed stimuli are related. Experiment 1 shows that sometimes related preexposure can be detrimental.

There are, of course, theories specifically designed to explain perceptual learning phenomena. The second author has written previously (McLaren et al., 1994) on the relationship between the theory of perceptual learning set out here, and the work of E.J. Gibson (1969), presenting some significant problems for Gibson's account. Hall's review of perceptual learning (Hall, 1991) is broadly in support of the main theoretical constructs of the McLaren et al. (1989) model (although there is some debate over the details). Also relevant is Goldstone's recently published work on the effect of categorization on perceptual discrimination (Goldstone, 1994), which, in terms of theory, is a development of the concepts of acquired distinctiveness (Miller, 1948) and acquired equivalence (Miller & Dollard, 1941). Acquired distinctiveness is the increased discriminability of two stimuli which results from their being given different labels; acquired equivalence is the decreased discriminability resulting from their being given the same label. These concepts are not in conflict with the general argument made here, and they are perhaps best considered as additional processes that may act when subjects begin to attach labels to stimuli. Although the labels are generally provided by the experimenter on an item-by-item basis, there seems no reason in principle why labels that subjects create themselves should not also produce acquired distinctiveness or acquired equivalence effects. Whether such effects have an important influence on stimulus representation as free classification proceeds is an open question, and one not readily answerable from the data presented in this paper.

In summary, preexposure does not simply fine-tune the classifications that would have been formed anyway—it can dynamically alter the actual classifications formed. If we are to understand how a learner selects the categories to be formed from among all those that could be formed, the effect that perceptual learning has on this process must also be understood. Many have argued that categorization is central to cognition. If this is true, perceptual learning deserves an equally lofty status.

REFERENCES

Ahn, W-K., & Medin, D.L. (1992). A two-stage model of category construction. *Cognitive Science*, *16*, 81–121.

Aitken, M.R.F., Bennett, C.H., McLaren, I.P.L., & Mackintosh, N.J. (1996). Perceptual differentiation during categorisation learning by pigeons. *Journal of Experimental Psychology: Animal Behavior Processes*, *22*(1), 43–50.

Anderson, J.R. (1990). *The adaptive character of thought*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Anderson, J.R. (1991). A rational analysis of categorization, *Psychological Review*, *98*, 409–429.

Bersted, C.T., Brown, B.R., & Evans, S.H. (1969). Free sorting with stimuli in a multidimensional attribute space. *Perception & Psychophysics*, *6B*, 409–413.

Bruner, J.S., Goodnow, J.J., & Austin, G.A. (1956). *A study of thinking*. New York: Wiley.

Chamizo, V.D., & Mackintosh, N.J. (1989). Latent learning and latent inhibition in maze discriminations. *The Quarterly Journal of Experimental Psychology*, *41B*(1), 21–31.

Cramér, H. (1946). *Mathematical methods of statistics*. Prineton, NJ: Princeton University Press.

Diamond, R., & Carey, S. (1986). Why faces are and are not special: An effect of expertise. *Journal of Experimental Psychology: General*, *115*, 107–117.

Estes, W.K. (1994). *Classification and cognition*. Oxford: Oxford University Press.

Evans, S.H., & Arnoult, M.D. (1967). Schematic concept formation: Demonstration in a free sorting task. *Psychonomic Science*, *9*(4), 221–222.

Fisher, D. (1987). Knowledge acquisition via incremental conceptual clustering. *Machine Learning*, *2*, 139–172.
Gibson, E.J. (1969). *Principles of perceptual learning and development*. New York: Appleton-Century-Crofts.
Gibson, E.J., & Walk, R.W. (1956). The effect of prolonged exposure to visually presented patterns on learning to discriminate them. *Journal of Comparative and Physiological Psychology*, *49*, 239–242.
Goldstone, R. (1994). Influences of categorization on perceptual discrimination. *Journal of Experimental Psychology: General*, *123*(2), 178–200.
Goss, A.E. (1953). Transfer as a function of type and amount of preliminary experience with task stimuli. *Journal of Experimental Psychology*, *55*, 258–269.
Greenhouse, S.W., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika*, *24*, 95–112.
Hall, G. (1979). Exposure learning in young and adult laboratory rats. *Animal Behaviour*, *27*, 586–591.
Hall, G. (1991). *Perceptual and associative learning*. Oxford: Clarendon Press.
Howell, D.C. (1992). *Statistical methods for psychology* (3rd ed.). Belmont, CA: Duxbury Press.
Kruschke, J.K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review*, *99*, 22–44.
Lubow, R.E. (1989). *Latent inhibition and conditioned attention theory*. Cambridge: Cambridge University Press.
Mackintosh, N.J., Kaye, H., & Bennett, C.H. (1991). Perceptual learning in flavour aversion conditioning. *The Quarterly Journal of Experimental Psychology*, *43B*(3), 297–322.
McClelland, J.L., & Rumelhart, D.E. (1985). Distributed memory and the representation of general and specific information. *Journal of Experimental Psychology: General*, *114*(2), 159–188.
McLaren, I.P.L. (1997). Categorization and perceptual learning: An analogue of the face inversion effect. *The Quarterly Journal of Experimental Psychology*, *50A*(2), 257–273.
McLaren, I.P.L., Bennett, C.H., Guttman-Nahir, T., Kim, K., & Mackintosh, N.J. (1995). Prototype effects and peak shift in categorization. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *21*(3), 662–673.
McLaren, I.P.L., Kaye, H., & Mackintosh, N.J. (1989). An associative theory of the representation of stimuli: Applications to perceptual learning and latent inhibition. In R.G.M. Morris (Ed.), *Parallel distributed processing – implications for psychology and neurobiology*. Oxford: Oxford University Press.
McLaren, I.P.L., Leevers, H.L., & Mackintosh, N.J. (1994). Recognition, categorization and perceptual learning. In C. Umilta & M. Moscovitch (Eds.), *Attention & üperformance XV* Cambridge, MA: MIT Press.
Medin, D.L., & Schaffer, M.M. (1978). Context theory of classification learning. *Psychological Review*, *85*(3), 207–238.
Michalski, R.S., & Stepp, R.E. (1983). Learning from observation: Conceptual clustering. In R.S. Michalski, J.G. Carbonell, & T.M. Mitchell (Eds.), *Machine learning: An artificial intelligence approach*. Palo Alto, CA: Tioga.
Miller, N.E. (1948). Theory and experiment relating psychoanalytic displacements to stimulus–response generalization. *Journal of Abnormal and Social Psychology*, *43*, 155–178.
Miller, N.E., & Dollard, J. (1941). *Social learning and imitation*. New Haven, CT: Yale University Press.
Nosofsky, R.M. (1986). Attention, similarity and the identification–categorization relationship. *Journal of Experimental Psychology: General*, *115*(1), 39–57.
Nosofsky, R.M., Palmeri, T.J., & McKinley, S.C. (1994). Rule-plus-exception model of classification learning. *Psychological Review*, *101*(1), 53–79.
Rescorla, R.A., & Wagner, A.R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A.H. Black and W.F. Prokasy (Eds.), *Classical conditioning II: Current research and theory* (pp. 64–99). New York: Appleton-Century-Crofts.
Rosch, E. (1973). Natural Categories. *Cognitive Psychology*, *4*, 328–350.
Rosch, E., & Mervis, C.B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, *7*, 573–605.

Rosenthal, R., & Rosnow, R.L. (1984). *Essentials of behavioral research: Methods and data analysis*. New York: McGraw-Hill.

Rumelhart, D.E., Hinton, G.E., & Williams, R.J. (1986). Learning internal representations by error propagation. In D.E. Rumelhart, J.L. McClelland, & The PDP Research Group. *Parallel distributed processing: Explorations in the microstructure of cognition*. Cambridge, MA: MIT Press.

Rumelhart, D.E., & Zipser, D. (1986). Feature discovery by competitive learning. In D.E. Rumelhart, J.L. McClelland, &üThe PDP Research Group. *Parallel distributed processing: Explorations in the microstructure of cognition*. Cambridge, MA: MIT Press.

Shanks, D.R. (1995). *The psychology of associative learning*. Cambridge: Cambridge University Press.

Shepard, R. (1987, September). Towards a universal law of generalisation for psychological science. *Science*, 1317–1323.

Shepard, R.N., Hovland, C.L., & Jenkins, H.M. (1961). Learning and memorization of classifications. *Psychological Monographs*, *75*(13, Whole No. 517).

Smith, E.E., & Medin, D.L. (1981). *Categories and Concepts*. Cambridge, MA: Harvard University Press.

Tipper, S.P. (1985). The negative priming effect: Inhibitory priming by ignored objects. *The Quarterly Journal of Experimental Psychology*, *37A*, 571–590.

Wilcox, R.R. (1987). New designs in analysis of variance. *Annual Review of Psychology*, *38*, 29–60.

Wittgenstein, L. (1958). *Philosophical Investigations*. Oxford: Blackwell.

Yin, R.K. (1969). Looking at upside down faces. *Journal of Experimental Psychology*, *81*, 141–145.

Manuscript received 17 July 1997
Accepted version received 25 March 1998

Apprentissage perceptuel et classification libre

Deux expériences sont reportées sur les effets de pré-exposition sur la discrimination dans une tâche de classification libre. Dans cette tâche, les sujets doivent grouper des stimuli d'une façon qui semble raisonable. L'espérience 1 démontre que l'effet de pré-exposition dépend de la structure des stimuli. L'expérience 1b est la première démonstration d'un apprentissage retardé dû à une simple pré-exposition chez les sujets humains adultes (les démonstrations précédentes ont utilisé une pré-exposition accidentelle ou masquée). L'expérience 2 supporte les conclusions de l'expérience 1 et ajoute une démonstration de l'importance cruciale de la similarité des stimuli. Ensemble ces expériences éliminent des explications des phenomènes observés dans cette étude qui sont basées sur l'attention. Les expériences offrent de l'information nouvelle quant aux effets de la pré-exposition. La pré-exposition peut changer les classifications formées par les sujets en plus de changer la vitesse à laquelle elles sont construites. Les implications de ces données pour les théories de la formation des catégories et l'apprentissage sont considérées.

Aprendizaje perceptivo y clasificación libre

Se presentan dos experimentos que investigan los efectos de la preexposición del estímulo en la actuación discriminativa en una tarea de clasificación libre, usando humanos adultos como sujetos. En la clasificación libre a los sujetos se les pide que agrupen estímulos de alguna manera que a ellos les parezca razonable o sensata. El Experimento 1 muestra que el efecto de preexposición depende de la estructura del estímulo. El Experimento 1b es la primera demostración de un retraso en el aprendizaje a consecuencia de la simple preexposición en sujetos humanos adultos (las demostraciones previas han dependido de preexposición incidental o enmascarada). El Experimento 2 aporta más apoyo a las conclusiones del Experimento 1 y las amplia con la demostración de que la similitud del estímulo es un factor crucial. Considerándolos en conjunto, estos experimentos excluyen una clase de explicaciones basadas en la atención del fenómeno del que aquí se informa. Los experimentos también proporcionan información nueva sobre los efectos de la preexposición. La preexposición puede cambiar la forma actual de las clasificaciones de los sujetos además de alterar la razón a la que se forman. Se tienen en cuenta las implicaciones de estos resultados para las teorías actuales de la formación de categorias y el aprendizaje perceptivo.