

Why are artificial polymorphous concepts so hard for birds to learn?

Stephen E. G. Lea, A. J. Wills, and Catriona M. E. Ryan

University of Exeter, Exeter, UK

Stimulus sets defined in terms of artificial polymorphous concepts have frequently been used in experiments to investigate the mechanisms of discrimination of natural concepts, both in humans and in other animals. However, such stimulus sets are frequently difficult for either animals or humans to discriminate. Properties of artificial polymorphous stimulus sets that might explain this difficulty include the complexity of the individual stimuli, the unreliable reinforcement of individual positive features, attentional load, difficulties in discriminating some stimulus dimensions, memory load, and a lack of the correlation between features that characterizes natural concepts. An experiment using chickens as subjects and complex artificial visual stimulus sets investigated these hypotheses by training the birds in discriminations that were not polymorphous but did have some of the properties listed above. Discriminations that involved unreliable reinforcement or high attentional load were found to approach the difficulty of polymorphous concept discriminations, and these two factors together were sufficient to account for the entire difficulty. The usual kind of artificial polymorphous concept may not be a good model for natural concepts as they are perceived and discriminated by birds. A RULEX account of natural concept learning may be preferable.

One of the abilities that other animals share with humans is the capacity to discriminate the kind of ill-defined categories that occur in the real world and that correspond to many human natural language concepts. Herrnstein and Loveland (1964) first demonstrated that pigeons could discriminate such categories and, moreover, that they learned to do so surprisingly readily, considering the complexity of the stimulus sets involved in comparison with the stimuli usually

used in discrimination experiments with pigeons. Since then, a substantial experimental literature has investigated this capacity, particularly in birds (for review, see Huber, 2001). It is now recognized that, to avoid making assumptions about the animals' prior knowledge, in most cases this kind of performance should be described as concept discrimination rather than concept formation (Lea, 1984), and that the level of abstraction required for successful discrimination varies

Correspondence should be addressed to S. E. G. Lea, University of Exeter, School of Psychology, Washington Singer Laboratories, Exeter EX4 4QG, UK. Email: s.e.g.lea@exeter.ac.uk

This manuscript was prepared while Stephen Lea was a Visiting Scholar at the University of California, Berkeley. Facilities made available by the department and by Dr Lucy Jacobs are gratefully acknowledged, as are comments on the manuscript from Lucy Jacobs and Al Kamil. The planning of the experiments described here benefited from parallel experiments being carried out with human subjects with the support of Grant No. 9/S17109 to A. J. Wills from the UK Biotechnology and Biological Sciences Research Council. Earlier versions of this manuscript were presented at the meetings of the Psychonomics Society, Vancouver, November 2003, and the Comparative Cognition Society, Melbourne FL, USA, March 2004.

substantially between experiments (Herrnstein, 1990; Huber, 1999).

The recognized characteristic of natural language concepts is that there are no simple necessary or sufficient conditions for being an instance of the concept, and such concepts are therefore called polymorphous (Ryle, 1949). An obvious way of modelling such a concept is to construct a category on the basis of some number, n , of binary dimensions, and to define positive instances as those in which at least m dimensions take positive values; normally m is chosen to be greater than $n/2$. Such constructs are referred to as m -out-of- n artificial polymorphous concepts. They were used in experiments on human concept formation by Shepard, Hovland, and Jenkins (1961; it is their Type IV problem). They were first treated as an explicit model of natural polymorphous concepts by Dennis, Hampton, and Lea (1973) and Rosch and Mervis (1975). They have been used in experiments with birds by Lea and Harrison (1978) and many others since (e.g., Huber & Lenz, 1993, 1996; Jitsumori, 1993, 1996; Lea, Lohmann, & Ryan, 1993; Von Fersen & Lea, 1990).

If natural concept discriminations are often learned surprisingly readily, the surprising thing about artificial polymorphous concepts is that they are often difficult to learn. For example, humans find them more difficult to learn than either conjunctive or disjunctive concepts, even though they are a better model of natural categories than either concept (see Dennis et al., 1973). They are also difficult for birds to learn (e.g., Lea et al., 1993). This difficulty raises questions about the use of the artificial polymorphous concept as a model for natural categories. Correspondingly, it throws doubt on the popular multiple linear feature account of animals' natural concept discrimination (Aust & Huber, 2002; Lea & Ryan, 1990), since according to that account, the m -out-of- n artificial polymorphous concept should be a perfect model of a natural category.

The purpose of the present paper is to investigate the origins of the difficulty that birds have in learning to discriminate artificial polymorphous concepts, and thus to see whether the difficulty has

any consequences for the theory of natural category discrimination. There are a considerable number of possible hypotheses about the source of the difficulty, and none of them has been investigated in detail. The paper therefore seeks to determine which of them are the most promising candidates. The possibilities investigated here are:

Individual stimulus complexity. The need to use multiple dimensions results in complex, visually fussy stimuli that might be harder to discriminate than simple patches of colour or consistent patterns.

Unreliable reinforcement of features. It seems likely that early in the acquisition of an m -out-of- n discrimination, a subject will only have discriminated, or only have started to learn about, a subset of the relevant features, perhaps only a single feature. But the reinforcement for responding to any single feature will always be unreliable; for example, in a 3-out-of-5 concept such as that used by Lea et al. (1993), a subject that detected the positive value of a single dimension and always responded to stimuli on the basis of that feature alone would be reinforced only on 11/16 of trials. Partial reinforcement tends to reduce the rate of acquisition of visual discriminations (e.g., Sutherland, 1966; Williams, 1989), and in this case it will presumably slow the acquisition of discrimination of the most salient dimension, which might well be a necessary precursor to a more broadly based discrimination.

Attentional load. In order to respond correctly on every trial, the subject must take all n dimensions into account on each trial. It may be that this imposes an impossible attentional load. Early non-continuity theories of discrimination learning (e.g., Krechevsky, 1932) assumed that animals could only attend to one stimulus dimension at a time. This cannot be literally true, because animals can learn discriminations based on a conjunction of the experimenter-designated cues (e.g., Rescorla, 1972), and though they might in principle do so by elaborating single cues corresponding to these conjunctions, this kind of

explanation has been questioned even by those who put it forward (e.g., Cook, Riley, & Brown, 1992). In any case, conjunctive discriminations are often relatively difficult, and tasks that require attention to multiple dimensions are routinely found to be more difficult, or to require longer exposure, than corresponding single-dimension tasks (e.g., Blough, 1972; Cook, Cavoto, & Cavoto, 1996; Cook et al., 1992; Maki, Riley, & Leith, 1976).

Feature discrimination difficulties. In several experiments using m -out-of- n artificial polymorphous concepts, subjects have failed to come under the control of all the relevant features (e.g., Lea et al., 1993), or have required special training to bring them under such control (e.g., Von Fersen & Lea, 1990). If subjects are not able to access all the features, they will be unable to make a 100% correct categorization of stimuli.

Memory load. In addition to the need to attend perceptually to multiple dimensions on each trial, an m -out-of- n task requires the subject to learn and remember the valences of all the features. This imposes a memory load, which might make the overall task more difficult. Although pigeons are known to have a formidable memory for visual scenes (Vaughan & Greene, 1984), that memory is not unlimited (Cook, Levison, Gillett, & Blaisdell, in press).

Lack of feature correlation. Most previous experiments on artificial polymorphous concepts have used all possible variants of the stimuli during training and have used each of them equally often. This creates an unnatural universe of instances, in which every possible combination of features occurs, and the occurrence of the positive value of one dimension does not predict that any other dimension will take its positive value. As a result, most of the stimuli encountered are marginal cases. For example, with a 3-out-of-5 artificial polymorphous concept, 10 out of the 16 possible positive stimuli have only three positive features and thus differ from one of the negative stimuli by only a single feature. Arguably, natural

concepts exist precisely because in the natural world features are correlated, and humans develop concepts to describe clusters of features that tend to co-occur. If this is the case, then marginal instances will be encountered much less commonly in natural concepts than in experiments with artificial polymorphous concepts. Perhaps it is the lack of feature correlation, and the consequent prevalence of such marginal instances, that makes the artificial polymorphous concept difficult: Humans have been shown to be sensitive to the presence of feature correlation in artificial concept tasks (Anderson & Fincham, 1996).

The general research strategy was to devise learning tasks that were not polymorphous but did have the properties listed above, and to compare birds' rates of learning them with those of birds learning a standard artificial polymorphous concept discrimination. In principle, there are two questions that can be asked about each of these factors. First, we need to know whether the factor of interest has any effect on the rate of learning under concept-learning conditions, so that it might contribute to the difficulty of the polymorphous task. Second, we need to ask whether it makes the discrimination as difficult as a polymorphous discrimination, so that it might be the sole explanation of the difficulty. If none of the factors on its own appears to have a sufficient effect to account for the difficulty of the polymorphous task, we can then proceed to consider whether combinations of factors might do so.

Large differences between conditions were of interest, so relatively small group sizes were used. It was expected that some of the tasks would be difficult, so chickens rather than pigeons were used as subjects, because in some previous experiments they have been found to acquire difficult discriminations more quickly than pigeons (Ryan & Lea, 1994). The stimuli used were newly designed for this experiment. Like those used by Lea et al. (1993) and Von Fersen and Lea (1990), they involved five bipolar dimensions. Unlike those stimuli, but like the Brunswik faces used by Huber and Lenz (1993, 1996), the stimuli were constructed out of separable

dimensions—that is, it was possible to present each feature separately, with the remaining dimensions completely absent. But whereas in the Brunswick faces the dimensions were all variations on geometrical shape, in the present stimuli the five dimensions were designed to be as far as possible orthogonal in their likely effects on the birds' visual apparatus. Within those constraints, steps were taken to make the five dimensions as similar in salience as possible. Brightness, often a dominant cue (e.g., Lea & Ryan, 1983), was not used. On the basis of common laboratory experience and formal studies such as that of Huber, Troje, Loidolt, Aust, and Grass (2000), colour was expected to be a salient cue, so it was confined to a small area of the stimulus. Line orientation, which is also readily discriminated by pigeons (e.g., Bloomfield, 1967), was similarly used only in a small area. Texture, which Troje, Huber, Loidolt, Aust, and Fieder (1999) and Huber et al. (2000) found to be important in pigeon concept discrimination, was included in the controlled form of spatial frequency. Vertical intensity gradient, also found to be important by Huber et al., was included in the form of the inversion of a vertically asymmetric shape element. The most subtle dimension used was a difference between small shapes, and this was used on multiple objects in order to enhance its salience.

Method

Subjects

The subjects were 31 bantam chickens, 10 male and 21 female, obtained from agricultural markets. Before use they were habituated to handling and carrying so that they would not be stressed by transfer to and from the experimental apparatus. They were housed in groups of one male and two or three females, in indoor runs measuring approximately $1.5 \times 0.6 \times 0.5$ m. The birds were fed daily in their runs, after all experimental sessions were complete. Before each bird's experimental session its crop contents were examined by palpation, and any bird with food in its crop was not tested. They were also weighed

before and after all experimental sessions. At the completion of the experiment the birds were returned to agricultural use.

Apparatus

A total of 4 one-key operant chambers were used. Each consisted of a plywood box whose front wall was a 69×39 -cm aluminium panel. A video monitor was located 25 cm from this panel, outside the chamber. The chickens could view this screen through the transparent perspex response key (6.5×6.5 cm) positioned centrally on the front wall of the operant chamber, 26 cm above the floor. A shutter, operated by a rotary solenoid, could be used to prevent the bird from viewing the screen. The panel also carried a 7×7.5 -cm aperture, positioned 15 cm below the pecking key, giving access to a solenoid-operated food hopper containing a commercial mixture of wheat, barley, and maize. The availability of this food was signalled by a white light in the hopper aperture, which was operated when the hopper was presented. General illumination was given by a 3.5-W yellow-lensed houselight situated 12 cm above the pecking key. Masking noise was provided by a ventilation fan and by white noise relayed via a 35-ohm loudspeaker mounted on the back of the intelligence panel. A separate computer (PC-compatible, Pentium Pro) was used to generate the experimental stimuli for each test chamber. Both chambers and their stimulus generation computers were housed in a darkened testing room. A further PC-compatible microcomputer (Viglen 4DX266) running under the Windows 3.1 operating system was located outside this room and controlled and recorded all experimental events and responses, using a program written in Borland Delphi; a network link enabled this computer to instruct the computers attached to experimental boxes in the testing room to generate the appropriate stimuli. The chickens' behaviour during experimental sessions could be regularly monitored via digital cameras, fitted with a wide-angle lens, mounted on the right or rear wall of each chamber. Each chicken was assigned to a single test chamber for all stages of the experiment.

Stimuli

Stimuli were constructed using five bipolar dimensions. Figure 1 shows two of the stimuli that were used, illustrating the two values of all five dimensions. The two stimuli shown in Figure 1 were used as a basis for generating categories of stimuli for use in the experiment: The stimulus on the left was referred to as the platonic version (i.e., perfect exemplar) of Category A, and the stimulus on the right as the platonic version of Category B or the antiplatonic version of Category A. Category A consisted of all 16 stimuli that had at least three features in common with the platonic A stimulus shown in Figure 1, and Category B consisted of the remaining 16 stimuli. Figure 1 thus shows which feature on each of the dimensions was associated with Categories A and B respectively.

The dimensions were:

- *Stripe orientation*: The stripes superimposed on the central square could be either horizontal or vertical.
- *Background colour*: The central square could be either blue or yellow.
- *Field shape*: The shapes forming a field around the central square could be either stars or flower shaped.
- *Trapezium baseline*: The trapezium surrounding the field could be either upright

or inverted, giving it a long or short baseline, respectively

- *Flanker spatial frequency*: The border around the entire stimulus was made up of a pattern of randomly present or absent squares. These squares were either small or large and thus generated two different textures, involving, respectively, high and low spatial frequencies. All the high-frequency stimuli used the same pattern, as did all the low-frequency stimuli, but the two patterns were unrelated.

Apart from the blue or yellow backgrounds, all elements of the stimuli were coloured red. The three colours were at approximately the same brightness to the human eye. The background was black.

General procedure

The chickens were trained using conventional methods to find food in the hopper when it was raised, and to peck the key when the shutter opened, revealing a neutral grey display on the monitor. Reinforcement was then set at 3-s access to the food hopper, and the key peck requirement was raised to a fixed interval of 15 s. Once pecking was established on this schedule, the chickens were trained on a successive discrimination schedule. Trials were separated by an



Figure 1. Examples of the stimuli used in the experiments. Stimulus (a) was the perfect (platonic) exemplar of Stimulus Category A, and stimulus (b) was the perfect exemplar of Stimulus Category B. The stimuli were shown in red against a black background, except that the background to the central rectangle was yellow in stimulus (a) and blue in stimulus (b).

interval of 15 s to 45 s during which the shutter was closed. The shutter was then opened to reveal a stimulus, which was exposed for a period that varied randomly between 10 s and 15 s. At the end of this period, in positive trials, the shutter remained open until the next peck to the centre key, which led to a 3-s hopper operation; in negative trials, the shutter closed immediately. Key pecks during the first 10 s of stimulus exposure were counted and used to assess discrimination. Sessions consisted of 64, 72, or 80 trials, depending on the experimental condition; this variation was necessary in order to ensure within-session balance in stimulus use under different conditions. Half the trials involved positive stimuli, and half involved negative stimuli; positive and negative trials were given in a pseudorandom sequence, generated afresh for each session, constrained so that no more than three negative or three positive stimuli were presented consecutively. Performance in individual sessions was assessed by comparing the numbers of pecks in the first 10 s of positive and negative trials, using the proportional overlap statistic ρ introduced by Herrnstein, Loveland, and Cable (1976), which equals the probability that an arbitrarily chosen positive stimulus attracted a higher response rate than did an arbitrarily chosen negative stimulus. Peck rates on trials within a session that used the same stimuli were not averaged before calculating ρ ; if this procedure is used, the expected value of ρ is not affected by the number of trials per session, provided that the population distributions of the numbers of pecks to positive and negative trials are the same for sessions of different length. Sessions were given once per day, normally 5 days per week; the number of sessions given varied between training conditions as detailed under *Experimental design*, below.

Experimental design and procedure for particular conditions

Each subject was assigned to one of the following six training conditions:

Polymorphous. The 3-out-of-5 artificial polymorphous stimulus categories A and B generated from

the two stimuli shown in Figure 1 were used. All 32 variants of the stimuli were used. Six birds were trained under this condition; for three of them, the Category A stimuli were designated as positive, and the Category B stimuli were designated as negative, while for the remaining three birds, these contingencies were reversed. Sessions consisted of 64 trials, and all 32 stimuli were used once in each half of the session. Three birds were trained under these conditions for a total of 20 sessions, and two for 60 sessions; the sixth bird was trained for 30 sessions but then died.

Platonic. Only the two stimuli shown in Figure 1 were used. These are the perfect exemplars of the two categories used under polymorphous conditions—hence the title “platonic”. Sessions consisted of 64 trials. Four birds were trained under these conditions until they reached a criterion of a ρ value equal or greater than .80 in two successive sessions; this required three sessions for three of the birds and four for the remaining one. For two of the birds, the stimulus shown on the left of Figure 1 was positive, and the stimulus on the right was negative; for the other two birds these contingencies were reversed.

Platonic-unreliable. As in the platonic condition, only the two stimuli shown in Figure 1 were used, and sessions consisted of 64 trials. However, within each of two blocks of 32 trials, reinforcement contingencies were normal for only 11 of the presentations of each of the two stimuli and were reversed for the remaining 5, to match the probability of reinforcement of a single feature under polymorphous conditions. Six birds were trained under these conditions for 20 sessions each; for three of them the stimulus shown on the left of Figure 1 was usually positive, and the stimulus on the right was usually negative; for the other three birds these contingencies were reversed.

Easy-polymorphous. Subsets of Categories A and B were used. They included only the 12 stimuli that had at least four features in common with one of the platonic stimuli shown in Figure 1

(and therefore no more than one feature in common with the other platonic stimulus). Sessions consisted of 72 trials, divided into six 12-trial blocks within which each of the 12 stimuli was used once. Six birds were trained for 10 sessions. For three birds, the stimuli from Category A were designated as positive, and the stimuli from Category B were designated as negative; for the remaining three birds, these contingencies were reversed.

Features in series. The five individual elements of the two stimuli shown in Figure 1 were used. Five birds were trained, for up to 50 sessions of 64 trials, in five blocks of up to 10 sessions. In each block of sessions only the two stimuli showing the two opposite features on a single dimension were used, the elements involving the remaining dimension being absent. For example, in the session block involving the background colour dimension, the stimuli were a blue square and a yellow square. The order in which the dimensions were used was varied between birds in a Latin square design. Within each block of sessions, if a bird reached a criterion of two successive sessions with a ρ value of at least .8, training was stopped, and the bird was moved on to the next block. For two of the birds, the elements of the platonic A stimulus were designated as positive, and the elements of the platonic B stimulus were designated as negative; for the remaining three birds, these contingencies were reversed.

Features in parallel. As in the features in series condition, the five individual elements of the two stimuli shown in Figure 1 were used. On any trial, any of the 10 possible stimuli could occur. Sessions consisted of 80 trials, divided into eight blocks of 10 trials within which each stimulus was used once. Training was continued for 40 sessions. Four birds were trained in this condition; for two of them, the elements of the platonic A stimulus were designated as positive, and the elements of the platonic B stimulus were designated as negative; for the remaining two birds, these contingencies were reversed.

Results

Results were examined for differences as a function of sex of the subject, or the chamber that it was tested in, and no effects were noted. These variables are therefore excluded from subsequent analyses.

The various hypotheses for the origin of the difficulty of polymorphous discriminations were examined in a series of planned tests. In some of these tests, we are able to make a prediction that should hold if a factor plays any role at all in the difficulty of the polymorphous condition; in others we are only able to predict what would happen if the factor is the sole source of that difficulty.

Individual stimulus complexity. If the only problem with the polymorphous condition is the complexity of the individual stimuli, performance in that condition should not differ from performance in the platonic condition, since the individual stimuli used in that condition are equally complex. Figure 2 shows acquisition under polymorphous and platonic conditions. For the platonic group, only data for the first three sessions are shown, since by this time three of the four birds in that condition had reached the criterion of ρ equal to or greater than .80 in two successive sessions; the fourth bird reached the criterion after one further session. None of the birds in the polymorphous condition reached the criterion within 20 sessions. The number of sessions to reach criterion differed significantly for the two groups: two-tailed Mann-Whitney test, $U(4, 6) = 0, p < .01$.

Unreliable reinforcement of features. If the only problem with the polymorphous condition is the unreliability of reinforcement to each feature, then performance in the platonic-unreliable condition should be worse than that in the platonic condition, but should not differ from performance in the polymorphous condition, since the reinforcement contingencies associated with each feature are equally unreliable in that condition. Figure 2 compares acquisition under these three conditions, over the first 20 sessions of training. In the analysis of the data under platonic-unreliable conditions, the response rates to trials on which the platonic A

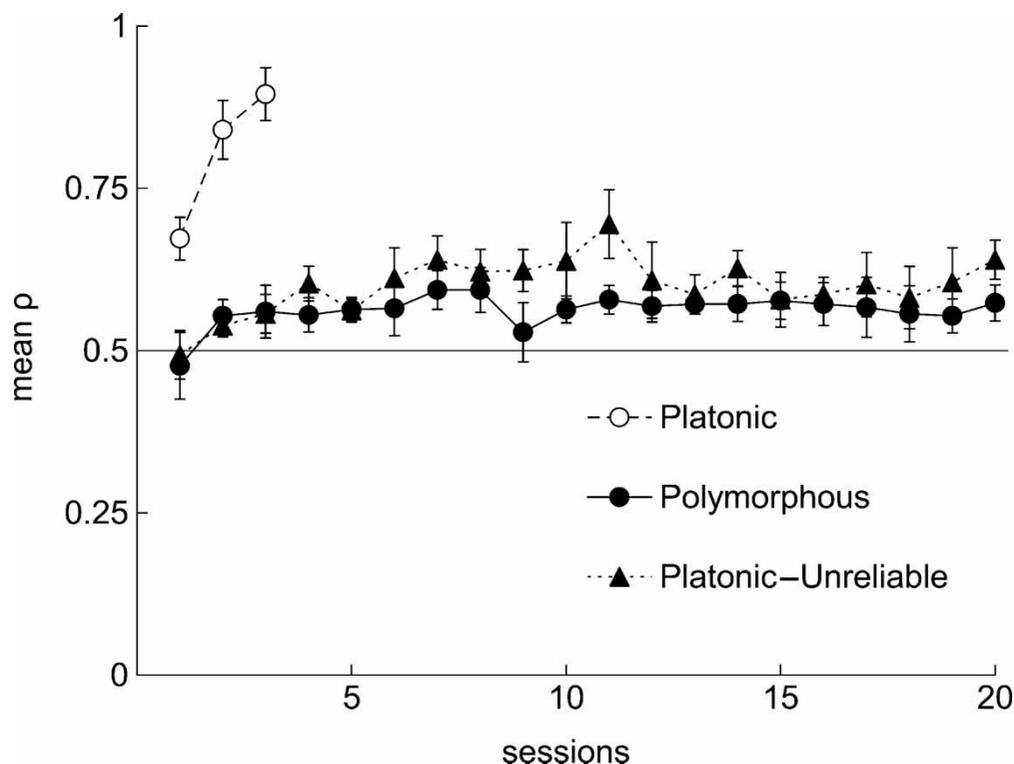


Figure 2. Acquisition in groups of birds trained under platonic, polymorphous, and platonic-unreliable conditions. Data are shown as means plus or minus standard errors.

and platonic B stimuli were presented were compared; in 5/16 of these, the reinforcement contingencies normally associated with those stimuli were reversed. Performance of the birds in the platonic and platonic-unreliable groups was compared in terms of the number of sessions required to reach the criterion of two consecutive sessions with ρ values of .8 or above. None of the birds in the platonic-unreliable group reached this criterion within 20 sessions, whereas all of the birds in the platonic group did so within three or four sessions; the rates of acquisition of these two groups therefore differed significantly: two-tailed Mann-Whitney test, $U(4, 6) = 0$, $p < .01$. Because none of the birds in the platonic-unreliable or the polymorphous group reached the criterion within the number of sessions given, their acquisition rates were compared in terms of the ρ values shown in Figure 2 for these two groups. For consistency

with later analyses, sessions were grouped into blocks of five containing 400 trials. An analysis of variance (ANOVA) with group and session block as factors revealed a significant interaction between groups and session block, $F(3, 30) = 2.99$, $p < .05$; the main effect of groups fell just short of significance, $F(1, 10) = 4.55$, $p = .06$, but the simple main effect of group was significant for the last session block, $t(10) = 2.74$, $p < .05$.

Attentional load. If a bird trained under polymorphous conditions is unable to attend to more than one stimulus dimension on a given trial, then even if it discriminates all features perfectly, its performance would be bounded. The positive value of each dimension is only present in 11 out of the 16 positive stimuli. So, regardless of which dimension it is attending to on a given trial, a subject that can only attend to one stimulus

dimension only has an 11/16 chance of making a correct response, even if it discriminates that feature (or, indeed, all five features) perfectly. Similarly, if a bird in the easy-polymorphous condition only attends to a single dimension, it can only categorize the stimulus correctly on 5/6 trials. It can be shown that these bounds correspond to bounds on ρ of .69 and .83 in polymorphous and easy-polymorphous conditions, respectively. Figure 3 compares performance on the polymorphous condition (twenty 80-trial sessions) and the easy-polymorphous condition (twelve 72-trial sessions) with these bounds. It is clear that they are not exceeded, at least with this amount of training. The mean ρ values for each bird over Sessions 16–20 (polymorphous conditions) and Sessions 6–10 (easy-polymorphous conditions) were compared with the hypothetical bound for each bird. In each group, all six birds had mean terminal ρ values below the hypothetical bound; the ranges were .51 to .64 for the polymorphous group, and .71 to .79 for the easy-polymorphous group. The proportion of birds falling below the hypothetical bound was significant in each group ($p < .05$, two-tailed binomial test). The two birds in the polymorphous group that were trained for a

further 40 sessions did not reach the hypothetical bound of .69: Their mean ρ values in sessions 55–60 were .60 and .67.

Feature discrimination difficulties. Figure 4 shows performance on each of the five dimensions of the

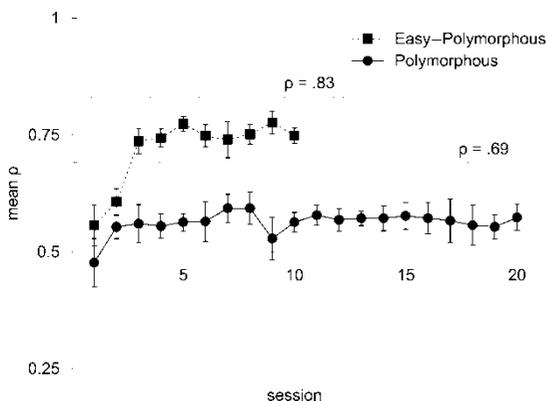


Figure 3. Acquisition under the easy-polymorphous and polymorphous conditions compared with the hypothetical bounds on performance under the multidimensionality hypothesis. Data are shown as means plus or minus standard errors.

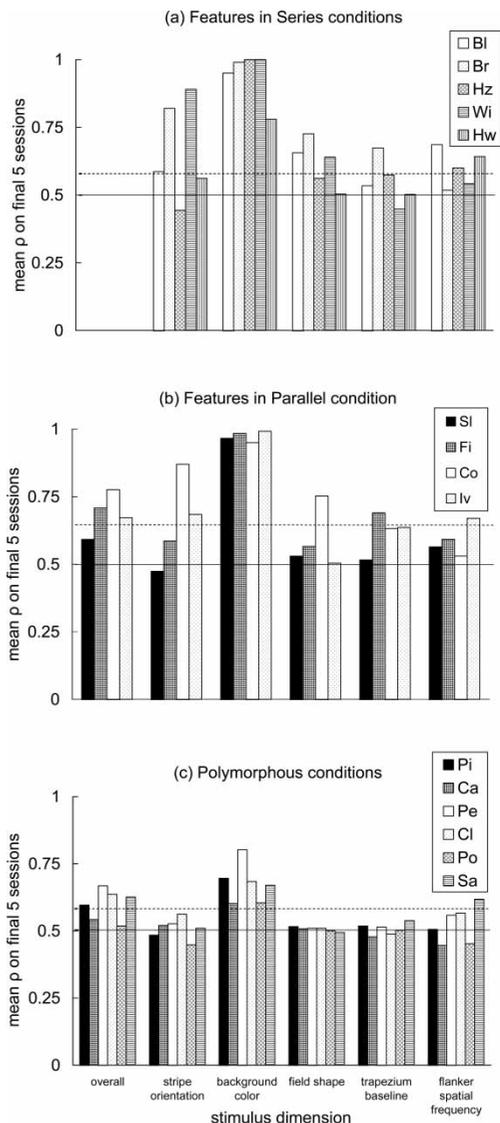


Figure 4. Performance (mean ρ) by individual birds on individual dimensions in the last five sessions of training under features in series, features in parallel, and polymorphous conditions. Dashed horizontal lines show critical mean ρ values for significant discrimination.

individual birds trained under features in series, features in parallel, and polymorphous conditions. The graphs show the means of ρ values for each dimension separately, over the final five sessions of training under each condition. For birds in the series condition that reached the criterion of two consecutive sessions with ρ at least .8 in fewer than 10 sessions, the ρ value in the final session given was used for remaining sessions in forming the mean. For the features in parallel condition, a ρ for each dimension was calculated by considering only the trials when that dimension was presented. For birds trained under polymorphous conditions, ρ was calculated between stimulus sets defined by the positive or negative values of each dimension in turn. Note that for this group, perfect discrimination would result in ρ values of .69 for each dimension, because under polymorphous conditions, positive feature values sometimes occur in negative stimuli. With imperfect discrimination, performance better than .69 on a dimension is possible, but only if performance is worse than that level on other dimensions; for example, a bird that came under complete control of a single feature would obtain a ρ of 1.00 for that dimension and .50 for all others. Because ρ is directly related to the Mann-Whitney U statistic, it is possible to calculate critical values of ρ that indicate statistically significant discrimination, taking into account the number of trials in each session, the number of sessions averaged, and the number of dimensions involved (and hence the risk of spurious significance through repeated tests). The critical values are .58 for the features in series and polymorphous conditions, and .65 for the features in parallel condition. No bird in any group showed substantial or significant discrimination of all five dimensions. However, a number of birds showed significant discrimination of more than one dimension (all birds in the features in series group, all birds except S1 in the features in parallel group, and bird Sa in the polymorphous group), and every dimension was discriminated significantly by at least two of the birds. In all three conditions the colour dimension was discriminated significantly, and better than any other dimension, by every bird, but there is

little consistency about which of the remaining dimensions were better discriminated.

Memory load. The memory load hypothesis states that the difficulty of polymorphous discrimination arises because birds have difficulty remembering the contingencies associated with five different features. If memory load plays any role at all, we can predict that (a) the features in parallel condition, in which any dimension may occur in any trial, should lead to slower learning than the features in series condition, in which training is given on each dimension in turn. We might also predict that (b) the five distinct dimension discriminations trained under features in series conditions will lead to successively slower learning, since the bird is already remembering more and more previous training, though this prediction may depend on assumptions about forgetting rates. Finally, if memory load is the sole explanation for the difficulty of learning under polymorphous conditions, we can predict that (c) learning in the features in parallel condition should be as difficult as learning in the polymorphous condition, since both conditions involve remembering the same five feature discriminations. Figure 5 tests prediction (a), by comparing overall acquisition under features in series and features in parallel conditions, using as independent variable the mean ρ in 400-trial blocks because of the different

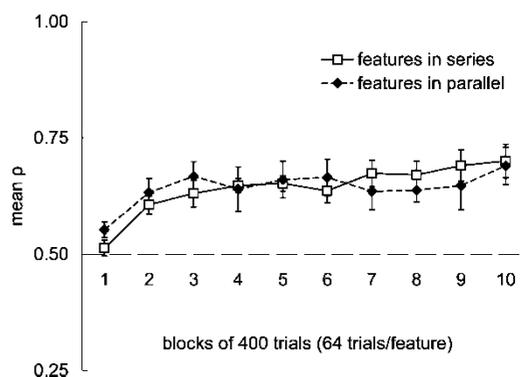


Figure 5. Performance of birds trained under features in series and features in parallel conditions, in successive blocks of 400 trials (each containing 64 training trials with each of the five dimensions). Data are shown as mean ρ values plus or minus standard errors.

numbers of trials per sessions required in the two conditions. There is no evidence that features in parallel training leads to slower acquisition: ANOVA showed that neither the main effect of groups, $F(1, 7) = 0.00$, $p = .98$, nor its interaction with session blocks, $F(9, 63) = 1.45$, $p = .19$, approached significance. Figure 6 tests prediction (b), by showing ρ values averaged over all 10 sessions of training with each dimension in the features in series condition, as a function of the order in which the birds were exposed to the five dimensions. Where birds in the series condition reached the criterion of two consecutive sessions with ρ at least .8 in under 10 sessions, the mean of the ρ value in the final session given was used for remaining sessions in forming a mean. It is clear that there is no tendency for slower acquisition of successive discriminations under series conditions; if anything, there is a slight trend towards quicker acquisition of later tasks, though this trend is not significant (Spearman's ρ for the mean across birds = .70, $p > .10$). Figure 7 tests prediction (c), by comparing acquisition in the

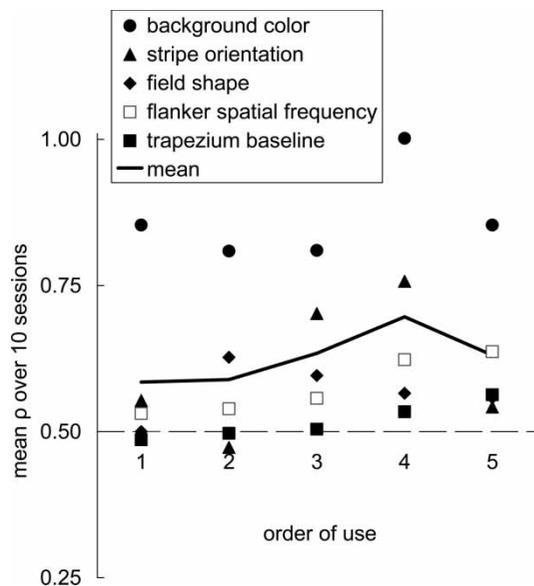


Figure 6. Mean ρ for each dimension under features in series conditions, over all sessions of training, as a function of the order in which the dimensions were used. Each point plotted represents data from a single bird.

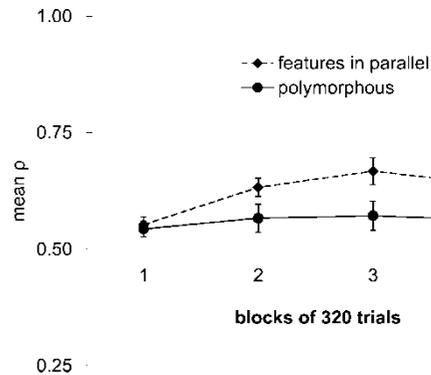


Figure 7. Performance of birds trained under features in parallel and polymorphous conditions, in successive blocks of 320 trials. Data are shown as mean ρ values plus or minus standard errors.

features in parallel and polymorphous conditions. Since session lengths were different for these two conditions, data are shown plotted against successive 320-trial blocks (four sessions for the features in parallel group and five for the polymorphous group). Data are shown only for the first four such blocks, corresponding to the 20 sessions experienced by the polymorphous birds. There is a clear difference between the two groups; the main effect of groups fell just short of significance in an ANOVA, $F(1, 8) = 3.91$, $p = .08$, but the interaction between groups and session block was significant, $F(3, 24) = 3.04$, $p < .05$, and the nature of the interaction, evident from Figure 7, was confirmed by testing the simple main effects of groups; this effect was significant for the third session block only, $t(8) = 3.00$, $p < .05$.

Feature correlation. If the difficulty of polymorphous discriminations arises entirely because the positive features are not correlated within a polymorphous stimulus set, and there is a consequent predominance of marginal instances, then no difficulty should be found in the easy-polymorphous condition, since with the restricted stimulus sets used in that condition, all features are correlated (each has a correlation of .33 with every other), and the marginal instances are not presented. That is, learning in the easy-polymorphous

group should be faster than that in the polymorphous group, in fact as fast as that in the platonic group. Performance of the easy-polymorphous and polymorphous groups can be compared in Figure 3. The easy-polymorphous condition evidently did lead to substantially faster learning than did the polymorphous condition. ANOVA confirmed this: The number of sessions was too low to allow the creation of blocks of equal numbers of trials, but using sessions as a factor, the main effect of groups was significant, $F(1, 10) = 43.62, p < .0005$; the test is conservative, because the sessions in the easy-polymorphous condition consisted of fewer trials than did those in the polymorphous condition. However, the easy-polymorphous discrimination was not as easy as the platonic discrimination (compare Figure 2). All four birds in the platonic group reached the criterion of two consecutive sessions with a ρ value of at least .8 in fewer trials than did any of the four birds in the easy-polymorphous condition, and this difference in the number of trials to criterion is significant: two-tailed Mann-Whitney test, $U(4, 4) = 0, p < .05$.

Discussion

Conclusions from tests of individual hypotheses

Summarizing the results of the individual tests, it has been shown that:

1. Individual stimulus complexity and memory load have either no effect or an effect that is not on the right scale to account for the difficulty of learning under polymorphous conditions.

2. The difficulty of the polymorphous discrimination cannot be solely due to the total absence of feature correlation and the consequent predominance of marginal exemplars, since introducing some feature correlation (in the easy-polymorphous condition) did not abolish the difficulty. It did make the discrimination somewhat easier, but there is an alternative explanation for that—the manipulation also reduced the unreliability of the reinforcement of individual features. The difference between the easy-polymorphous and polymorphous conditions could still be attributed to the effects of feature correlation, which is greater

in polymorphous than in easy-polymorphous conditions, even though it is present in both. But that difference would still fall well short of what is needed to explain the difficulty of the polymorphous conditions.

3. As in a number of previous experiments using m -out-of- n artificial polymorphous concepts (e.g., Lea et al., 1993; but contrast Huber & Lenz, 1993, 1996), the difficulty of discriminating the different dimensions varied sharply. Figure 4 establishes that all the dimensions could be discriminated by chickens under the present experimental conditions and within the amounts of training given here, but it was still the case that no bird gave evidence of discriminating all the dimensions. This was true even under conditions (features in series) where the birds were not required to respond to more than one stimulus dimension within a single block of sessions. One dimension, colour, was discriminated significantly by every bird in every condition. It is clear that in many experiments, including the present one, part of the difficulty of training birds in artificial polymorphous concept discriminations results from the practical difficulty of finding multiple dimensions all of which are readily discriminated by the subjects. However, in itself unequal discriminability is not a reason why the birds should not come under the control, after sufficient training, of all the dimensions that they can perceptually discriminate. For example, the theory of Rescorla and Wagner (1972), and its many derivatives, would all predict that all discriminable dimensions would eventually acquire control so long as they all contribute some independent information about the probability of reinforcement (as is necessarily the case with artificial polymorphous concepts). It is only in the presence of limited attentional capacity that unequal discriminability can affect the asymptotic breadth of stimulus control.

4. Two factors, attentional load and unreliable reinforcement of features, had effects that were of the right order of magnitude to account for the difficulty. However, on its own, neither is adequate as an explanation. The average

performance of the birds in the polymorphous and the easy-polymorphous groups was consistently and significantly worse than the bounds suggested by correct responding to a single feature on each trial; according to the attentional load hypothesis, these bounds should have been reached. It is of course possible that with more extended training the bounds would have been reached or exceeded. In particular, the easy polymorphous condition was included in the experiment primarily for a comparison with the platonic condition required by the feature correlation hypothesis, so the number of training sessions given was low. However, Figure 3 does not suggest that performance was tending towards the hypothetical bound under either condition, and the two birds in the polymorphous group that were trained for longer did not reach the bound even after 60 sessions. In any case, the data are sufficient to show that at least early in training, unreliability of feature reinforcement does not account for all the difficulty in acquiring polymorphous discriminations. With regard to the unreliable reinforcement of features hypothesis, although the unreliable-platonic conditions led to much poorer performance than did platonic conditions, there was still a significant difference between birds trained under these conditions and those trained under polymorphous conditions.

5. The fact that the birds trained under polymorphous conditions did not come under control of all five dimensions does not provide a trivial alternative to either of these two hypotheses. It is true that if the birds were only able to see one of the features in the stimuli, then their performance would be bounded at the ρ values shown in Figure 3 for reasons that have nothing to do with attention. But Figures 4a and 4b show that when the features were presented in isolation, all the dimensions were discriminated by at least some birds, and most birds discriminated most dimensions to at least some extent. It follows that the fact that under polymorphous conditions most features are not discriminated is not an explanation of the difficulty of polymorphous discriminations, but part of what we are trying to explain.

A synthetic approach

If none of these effects can explain the difficulty of the polymorphous discrimination on its own, is it possible that two or more of them in combination would do so? The obvious combination to investigate involves the two strong effects, unreliability of feature reinforcement and attentional load. According to the attentional load hypothesis, on any given trial a bird can only attend to a single dimension of the stimulus. We know that its maximum performance under polymorphous conditions will then be bounded, because the positive value of an arbitrarily chosen dimension will only be in the positive stimulus on a fraction of trials. In addition, however, according to the reinforcement unreliability hypothesis, the bird's learning will be slowed by the consequent unreliability of reinforcement. Can we estimate the combined effects of these two factors?

As it happens, we can. Figure 2 compares the performance of the polymorphous group with that of the platonic-unreliable group and shows that the polymorphous group performs worse. In that figure, performance of the platonic-unreliable group was assessed in terms of the stimulus shown on each trial: A response to the positive stimulus was counted as correct, even if the trial did not end in reinforcement. But under polymorphous conditions, we only count a response as correct if the trial ends in reinforcement. According to the combined unreliability of feature reinforcement and attentional load hypothesis, the features of the stimulus that are controlling the bird's behaviour under polymorphous conditions will not be positive on all such trials. There is no way of determining for sure what those features are on a trial-by-trial basis. However, we can put the two groups onto an equal footing by looking at the capacity of the platonic-unreliable group to discriminate trials that terminated in reinforcement, rather than trials in which the positive stimulus was shown. Figure 8 compares the performance of the polymorphous group with that of the platonic-unreliable group assessed on this alternative basis. It can be seen that performance was very similar under these two conditions, and ANOVA showed no suggestion of a significant

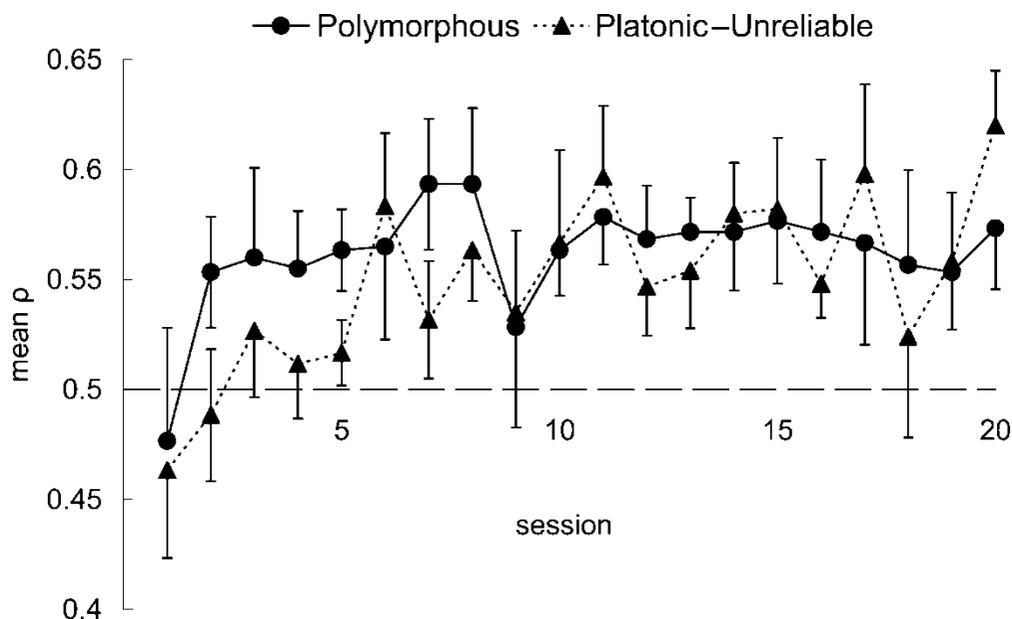


Figure 8. Comparison of performance of the polymorphous group with that of the platonic-unreliable group, assessing the performance of both groups by comparing responding on trials on which it was reinforced with responding on trials on which it was not reinforced. In the case of the platonic-unreliable group, the negative stimulus was presented on some reinforced trials and the positive stimulus on some nonreinforced trials. Data are shown as means plus or minus standard errors; error bars are not drawn where they would overlap.

difference either as a main effect of groups, $F(1, 10) = 0.26$, $p = .62$, or as an interaction between groups and session blocks, $F(19, 190) = 0.57$, $p = .92$.

Implications

Although the present experiments used artificial polymorphous stimulus sets, their goal, like that of all such experiments, is to help us to understand how animals discriminate natural stimuli. On the basis of the present data, three conclusions can be drawn about natural concept discriminations.

First, confirming what has already been found in numerous experiments, it is difficult and sometimes impossible to find multiple artificial stimulus dimensions that will be equally, or even approximately equally, discriminable and salient for birds. It seems unlikely, therefore, that there will be many such dimensions in sets of natural photographs, chosen for their membership of categories defined in terms of an experimenter's concepts.

Consistent with this is the finding in many experiments with natural stimuli, that relatively few, relatively simple dimensions give a good account of birds' discrimination of such stimulus sets (e.g., luminous flux, Lea & Ryan, 1983; texture, Huber et al., 2000; eyes and eyebrows in cartoon faces, Matsukawa, Inoue, & Jitsumori, 2004).

Second, the results are consistent with the view that birds have difficulty in distributing attention between numerous features on individual trials. This conclusion too is supported by much previous literature, including the recurrent finding of relatively slow learning of conjunctive or configural discrimination tasks, which require simultaneous attention to multiple dimensions or the elaboration of compound cues from the elements presented (Cook et al., 1992; Rescorla, 1972; for theoretical treatment, see Pearce, 1994).

Finally, the unreliable reinforcement associated with any single feature in a truly polymorphous stimulus category will profoundly slow down

learning. This too is consistent with the slow acquisition of configural discriminations.

However, if a stimulus set defined in terms of a natural category is not well modelled by an artificial polymorphous concept, how can we model it? How are we to explain birds' capacity to discriminate accurately between stimulus sets for which no single feature provides a sufficient criterion? Do we have to fall back on birds' excellent performance in absolute discrimination tasks (Cook et al., in press; Vaughan & Greene, 1984; Von Fersen & Delius, 1989) and conclude that birds do not learn concept discrimination tasks on any kind of conceptual basis at all?

The frequent finding of slow learning of pseudoconcept control tasks shows that exclusive reliance on absolute discrimination would be an inadequate account of animal concept discrimination. However, it may be possible to combine an absolute discrimination approach with a single-dimension (or few-dimension) learning approach. What distinguishes artificial polymorphous concepts from natural ones may be the fact that in an artificial concept, each of the features is equally valid. In many natural concept discriminations, a single feature or a localized feature cluster would provide a sufficient basis for discriminating the majority of the stimuli—certainly a higher proportion than the 11/16 that a single feature classifies correctly in a 3-out-of-5 polymorphous task. Although as noted above intermittent reinforcement retards discrimination learning, it is certainly no bar to the acquisition of highly reliable discrimination (as is shown, for example, by the success of multiple schedules of reinforcement, see Ferster & Skinner, 1957, chap. 10). The first consequence of such a dominant feature is therefore likely to be that all stimuli are responded to in terms of a single dimension. However, it is plausible to assume that if a bird identifies a single dominant dimension and comes to use that to classify most of the stimuli, it would then be able to learn to categorize the few exceptional cases by absolute discrimination. It is notable that human participants learning artificial polymorphous discriminations (e.g., Dennis et al., 1973) often comment that they worked in

terms of a simple feature rule plus some exceptions, and several modern models of human concept learning incorporate an explicit rule plus exceptions (RULEX) principle (e.g., Anderson & Betz, 2001; Nosofsky, Palmeri, & McKinley, 1994).

A RULEX hypothesis may not provide a complete account of the present data. The easy-polymorphous stimulus sets had very much the kind of structure that should be easily learned by a RULEX system: A single feature would classify 10/12 of the stimuli, leaving only two exceptional cases to be learned by absolute discrimination. Yet, as Figure 3 shows, the birds' performance with these stimulus sets showed no signs of approaching perfect discrimination. However the easy-polymorphous sets differed in two ways from the hypothesized structure of natural concepts. First, the five dimensions were equally valid, so there was no "key" feature to provide an easy route for initial learning. Second, to the human eye at least, the individual stimuli were highly confusable, much more so than individual instances of a natural concepts are likely to be.

Numerous studies have examined the features correlated with pigeons' response rates to different stimuli in concept discrimination tasks, both between and within categories. They have consistently shown that, under a linear feature model, multiple features must be invoked to predict performance accurately, both in training conditions and in transfer tests (e.g., Aust & Huber, 2002; Jitsumori & Yoshihara, 1997; Lea & Ryan, 1983; Lubow, 1974). However, this result does not necessarily imply that all those features are abstracted by the pigeons, or that linear combination is the mechanism by which information from different dimensions is integrated. A correlation of multiple feature values with response rate would also be generated if the birds were responding to a single feature but in addition discriminating a limited number of exceptional stimuli absolutely. Such a mechanism can only give perfect performance with the kind of finite stimulus set used in the present experiment, which contrasts with the effectively infinite sets used in some studies (e.g., Herrnstein & Loveland, 1964, or see Aust & Huber, 2003).

But it can still generate good overall performance even if individual stimuli were rarely or never repeated, so long as the proportion of exceptions to the simple rule is not too high. Moreover, a RULEX approach is consistent with the results of some recent experiments on avian concept discrimination. For example, Aust and Huber (2001) revisited the original Herrnstein and Loveland (1964) people-present/people-absent concept discrimination task, and they showed that pigeons that learned it successfully were responding to both category-specific and item-specific information. Item-specific information can be overridden if contingencies of reinforcement subsequently change (Loidolt, Aust, Meran, & Huber, 2003), and this makes it possible that with extensive training, there might be a transfer from control by a few features plus exceptions to a more truly multiple-feature control.

Original manuscript received 2 August 2004
Accepted revision received 9 February 2005
PrEview proof published online 13 July 2005

REFERENCES

- Anderson, J. R., & Betz, J. (2001). A hybrid model of categorization. *Psychonomic Bulletin and Review*, *8*, 629–647.
- Anderson, J. R., & Fincham, J. M. (1996). Categorization and sensitivity to correlation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 259–277.
- Aust, U., & Huber, L. (2001). The role of item- and category-specific information in the discrimination of people versus nonpeople images by pigeons. *Animal Learning & Behavior*, *29*, 107–119.
- Aust, U., & Huber, L. (2002). Target-defining features in a “people-present/people-absent” discrimination task by pigeons. *Animal Learning & Behavior*, *30*, 165–176.
- Aust, U., & Huber, L. (2003). Elemental versus configural perception in a people-present/people-absent discrimination task by pigeons. *Learning & Behavior*, *31*, 213–224.
- Bloomfield, T. M. (1967). A peak shift on a line-tilt continuum. *Journal of the Experimental Analysis of Behavior*, *10*, 361–366.
- Blough, D. S. (1972). Recognition by the pigeon of stimuli varying in two dimensions. *Journal of the Experimental Analysis of Behavior*, *18*, 345–367.
- Cook, R. G., Cavoto, K. K., & Cavoto, B. R. (1996). Mechanisms of multidimensional grouping, fusion, and search in avian texture discrimination. *Animal Learning & Behavior*, *24*, 150–167.
- Cook, R. G., Levison, D. G., Gillett, S., & Blaisdell, A. P. (in press). Capacity and limits of associative memory in pigeons. *Psychonomic Bulletin and Review*.
- Cook, R. G., Riley, D. A., & Brown, M. F. (1992). Spatial and configural factors in compound stimulus-processing by pigeons. *Animal Learning & Behavior*, *20*, 41–55.
- Dennis, I., Hampton, J., & Lea, S. (1973). New problem in concept formation. *Nature*, *243*, 101–102.
- Ferster, C. B., & Skinner, B. F. (1957). *Schedules of reinforcement*. New York: Appleton-Century-Crofts.
- Herrnstein, R. J. (1990). Levels of stimulus control: A functional approach. *Cognition*, *37*, 133–146.
- Herrnstein, R. J., & Loveland, D. H. (1964). Complex visual concept in the pigeon. *Science*, *146*, 549–551.
- Herrnstein, R. J., Loveland, D. H., & Cable, C. (1976). Natural concepts in pigeons. *Journal of Experimental Psychology: Animal Behavior Processes*, *2*, 285–302.
- Huber, L. (1999). Generic perception: Open-ended categorization of natural classes. *Cahiers de Psychologie Cognitive*, *18*, 845–887.
- Huber, L. (2001). Visual categorization in pigeons. In R. G. Cook (Ed.), *Avian visual cognition*. Retrieved 13 April 2004 from www.pigeon.psy.tufts.edu/avc/huber/
- Huber, L., & Lenz, R. (1993). A test of the linear feature model of polymorphous concept discrimination with pigeons. *Quarterly Journal of Experimental Psychology*, *46B*, 1–18.
- Huber, L., & Lenz, R. (1996). Categorization of prototypical stimulus classes by pigeons. *Quarterly Journal of Experimental Psychology*, *49B*, 111–133.
- Huber, L., Troje, N. F., Loidolt, M., Aust, U., & Grass, D. (2000). Natural categorization through multiple feature learning in pigeons. *Quarterly Journal of Experimental Psychology*, *53B*, 341–357.
- Jitsumori, M. (1993). Category discrimination of artificial polymorphous stimuli based on feature learning. *Journal of Experimental Psychology: Animal Behavior Processes*, *19*, 244–254.
- Jitsumori, M. (1996). A prototype effect and categorization of artificial polymorphous stimuli in pigeons.

- Journal of Experimental Psychology: Animal Behavior Processes*, 22, 405–419.
- Jitsumori, M., & Yoshihara, M. (1997). Categorical discrimination of human facial expressions by pigeons: A test of the linear feature model. *Quarterly Journal of Experimental Psychology*, 50B, 253–268.
- Krechevsky, I. (1932). Hypotheses in rats. *Psychological Review*, 39, 516–532.
- Lea, S. E. G. (1984). In what sense do pigeons learn concepts? In H. L. Roitblat, T. G. Bever, & H. S. Terrace (Eds.), *Animal cognition* (pp. 263–276). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Lea, S. E. G., & Harrison, S. N. (1978). Discrimination of polymorphous stimulus sets by pigeons. *Quarterly Journal of Experimental Psychology*, 30, 521–537.
- Lea, S. E. G., Lohmann, A., & Ryan, C. M. E. (1993). Discrimination of 5-dimensional stimuli by pigeons: Limitations of feature analysis. *Quarterly Journal of Experimental Psychology*, 46B, 19–42.
- Lea, S. E. G., & Ryan, C. M. E. (1983). Feature analysis of pigeons' acquisition of concept discrimination. In M. L. Commons, R. J. Herrnstein, & A. R. Wagner (Eds.), *Quantitative analyses of behavior: Vol 4. Acquisition processes* (pp. 239–253). Cambridge, MA: Ballinger.
- Lea, S. E. G., & Ryan, C. M. E. (1990). Unnatural concepts and the theory of concept discrimination in birds. In M. L. Commons, R. J. Herrnstein, S. Kosslyn, & D. Mumford (Eds.), *Quantitative analyses of behavior: Vol 8. Behavioral approaches to pattern recognition and concept formation* (pp. 165–185). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Loidolt, M., Aust, U., Meran, I., & Huber, L. (2003). Pigeons use item-specific and category-level information in the identification and categorization of human faces. *Journal of Experimental Psychology: Animal Behavior Processes*, 29, 261–276.
- Lubow, R. E. (1974). High-order concept formation in the pigeon. *Journal of the Experimental Analysis of Behavior*, 21, 475–483.
- Maki, W. S., Riley, D. A., & Leith, C. R. (1976). The role of test stimuli in matching to compound samples by pigeons. *Animal Learning and Behavior*, 4, 13–21.
- Matsukawa, A., Inoue, S., & Jitsumori, M. (2004). Pigeon's recognition of cartoons: Effects of fragmentation, scrambling, and deletion of elements. *Behavioural Processes*, 65, 25–34.
- Nosofsky, R. M., Palmeri, T. J., & McKinley, S. C. (1994). Rule-plus-exception model of classification learning. *Psychological Review*, 101, 53–79.
- Pearce, J. M. (1994). Similarity and discrimination: A selective review and a connectionist model. *Psychological Review*, 101, 587–607.
- Rescorla, R. A. (1972). "Configural" conditioning in discrete-trial bar pressing. *Journal of Comparative and Physiological Psychology*, 79, 307–317.
- Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current research and theory* (pp. 64–99). New York: Appleton-Century-Crofts.
- Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, 7, 573–605.
- Ryan, C. M. E., & Lea, S. E. G. (1994). Images of conspecifics as categories to be discriminated by pigeons and chickens—slides, video tapes, stuffed birds and live birds. *Behavioural Processes*, 33, 155–175.
- Ryle, G. (1949). *The concept of mind*. London: Hutchinson.
- Shepard, R. N., Hovland, C. I., & Jenkins, H. M. (1961). Learning and memorization of classifications. *Psychological Monographs*, 75 (517).
- Sutherland, N. S. (1966). Partial reinforcement and breadth of learning. *Quarterly Journal of Experimental Psychology*, 18, 289–301.
- Troje, N. F., Huber, L., Loidolt, M., Aust, U., & Fieder, M. (1999). Categorical learning in pigeons: The role of texture and shape in complex static stimuli. *Vision Research*, 39, 353–366.
- Vaughan, W., & Greene, S. L. (1984). Pigeon visual memory capacity. *Journal of Experimental Psychology: Animal Behavior Processes*, 10, 256–271.
- Von Fersen, L., & Delius, J. D. (1989). Long-term retention of many visual patterns by pigeons. *Ethology*, 82, 141–155.
- Von Fersen, L., & Lea, S. E. G. (1990). Category discrimination by pigeons using 5 polymorphous features. *Journal of the Experimental Analysis of Behavior*, 54, 69–84.
- Williams, B. A. (1989). Partial reinforcement effects on discrimination learning. *Animal Learning and Behavior*, 17, 418–432.