

Natural categorization through multiple feature learning in pigeons

Ludwig Huber

University of Vienna, Vienna, Austria

Nikolaus F. Troje

Ruhr University Bochum, Bochum, Germany

Michaela Loidolt, Ulrike Aust, and Dieter Grass

University of Vienna, Vienna, Austria

Recently (Troje, Huber, Loidolt, Aust, & Fieder 1999), we found that pigeons discriminated between large sets of photorealistic frontal images of human faces on the basis of sex. This ability was predominantly based on information contained in the visual texture of those images rather than in their configural properties. The pigeons could learn the distinction even when differences of shape and average intensity were completely removed. Here, we proved more specifically the pigeons' flexibility and efficiency to utilize the class-distinguishing information contained in complex natural classes. First, we used principal component as well as discriminant function analysis in order to determine which aspects of the male and female images could support successful categorization. We then conducted various tests involving systematic transformations and reduction of the feature content to examine whether or not the pigeons' categorization behaviour comes under the control of category-level feature dimensions—that is, those stimulus aspects that most accurately divide the stimulus classes into the experimenter-defined categories of “Male” and “Female”. Enhanced classification ability in the presence of impoverished test faces that varied only along one of the first three principal components provided evidence that the pigeons used these class-distinguishing stimulus aspects as a basis for generalization to new instances.

Requests for reprints should be sent to Ludwig Huber, Institute of Zoology, University of Vienna, Biocenter, Althanstrasse 14, A-1090 Vienna, Austria. Email: ludwig.huber@univie.ac.at

The research was supported by a grant from the Austrian Science Foundation (P 10975–BIO) to Ludwig Huber and a grant by the Alexander von Humboldt Foundation to Nikolaus F. Troje. We thank Michael Pollirer and Renate Nowotny for their help in the pigeon laboratory throughout this study and Chris Sturdy for correcting grammar and style of the manuscript.

Additional material and examples of the face stimuli found in this article can be found at Ludwig Huber's chapter in Robert G. Cook's (Ed.) cyberbook *Avian visual cognition*

(available at: <http://www.pigeon.psy.tufts.edu/avc/>)

and at Niko Troje's homepage (available at: <http://www.bio.psy.ruhr-uni-bochum.de/troje/>).

Researchers of visual categorization in pigeons have long been faced with a dilemma. Those who trained pigeons on complex stimulus classes, like photographs of natural scenes, could demonstrate the pigeons' capacity for categorization but lost the ability to identify the features to which the birds were responding. Those who used carefully constructed artificial stimuli, like simple line drawings, could study the underlying categorization mechanism but at the expense of a limited generalizability to natural visual classes. These traditions, called the *analytic* and the *synthetic* approach by Fersen and Lea (1990), can be viewed as alternative ways to test visual categorization in animals. Here, we report a study in which we sought to unify both approaches by using modern technical and statistical tools to discover which features pigeons use when solving a "naturalistic" classification task.

Recently, we started to investigate the visual classification by pigeons in a task that seemed to us an appropriate simulation of the categorization problem in nature (Troje, Huber, Loidolt, Aust, & Fieder, 1999). The design of that study and of the present one were developed by attending to the following aims:

- 1 The stimulus classes are natural categories that are consistent with natural stimulus variation and can be assumed to be evolved to be classified correctly. As representations of complex natural objects, the stimuli provide complex variations in both the spatial arrangements of stimulus components on the one hand and their particular appearance on the other hand. Specifically, the *surface properties* of objects, including the colour, intensity, and other first-order statistics (which we called "texture" in Troje et al., 1999) seem to be an essential part of the information that is contained in the visual array reflected by natural objects (Gibson, 1979).
- 2 Pigeons are naive with respect to that discrimination task at the onset of the experiment; neither innate nor individually acquired concepts should be likely to interfere with the categorization rule.
- 3 The object representation should provide an efficient coding scheme. Instead of the commonly used pixel-based representation of objects it is preferable to use a stimulus space in which the objects are coded by means of corresponding features (*correspondence-based representation*). Obtaining such sophisticated representations allows the use of simple distance measures and linear classifiers.

The first two requirements were satisfied by choosing human faces as stimuli (as did Jitsumori & Yoshihara, 1997, and Wasserman, Tassinari, Bhatt, & Sayasenh, 1989), and the polymorphous feature "sex" as class definition. The third goal was achieved some years ago by Vetter and Troje (1995) who developed a feature-based representation for images of faces by exploiting pixel-by-pixel correspondence between the images. The features used for establishing correspondence were not semantically predefined high-level features like "eye" or "nose", but were established on the single-pixel level by using adapted optical flow algorithms (Bergen, Anandan, Hanna, & Hingorani, 1992). One important advantage of this kind of object representation is that it makes it possible to separate the information contained in the images of human faces by addressing the spatial arrangements of stimulus features ("shape") on the one hand and their particular appearance ("visual texture") on the other hand (Troje & Vetter, 1998; Vetter & Troje, 1997).

The main result of our first experiments with these stimuli was the differential contribution of texture and shape to categorization decisions about the sex of human faces (Troje et al., 1999). Three groups of eight pigeons each were confronted with different versions of the same set of stimuli differing according to the amount of shape and texture information that they contained. Only the pigeons that were presented with texture information (with or without additional shape information) successfully sorted 200 faces into the categories “Male” and “Female”. Of those faces, 100 were associated with differential reinforcement during training, but the remaining 100 faces were spontaneously classified thereafter with no measurable difference in performance.

In further experiments we sought to identify the reason for the superiority of the texture task over the shape task. One experiment was designed to verify whether the pigeons used particular image attributes of those two feature domains, such as the average intensity or size of the faces. Within-class correlations between those attributes and the pecking rates during classification training indicated a predominant role for both of them. This result was confirmed by inserting test images with ambiguous size or overall intensity and measuring the degree of disruption on the pigeons' classification ability. We concluded that the surface properties of natural objects serve as a very general and potent perceptual dimension in the pigeon's visual environment.

A final experiment was conducted in order to determine whether the subjects were also able to learn the sex discrimination when the images were normalized for both their shape and for their overall intensity (the “texture-only, intensity-normalized” situation; subsequently labelled “normalized”). The positive results provided evidence that pigeons are able to successfully shift strategies and use other stimulus aspects for the appropriate division of the reduced stimulus space. Although not one of the Troje et al. (1999) study's original aims, which were mainly oriented at *perceptual* questions, these findings have important implications for the way in which pigeons use features to classify stimuli.

Of course, one is tempted to explain the preference for the texture task by claiming that the pigeons are able to extract some relevant information from the feature space by means of an acquired adjustment of the saliency of features contained in the surface domain. However, it is also conceivable that a naive pigeon approaches the classification task with a fixed set of ways of analysing pictures (the *feature analysis hypothesis*; see Lea, 1984). If this were the case, the difference between the texture and the shape group in Troje et al. (1999) would be a matter of perception rather than of cognition; their findings would be of little relevance for theories of categorization (see also Watanabe, Lea, & Dittrich, 1993).

A basic requirement for the claim that pigeons learn to use the relevant feature dimension rather than coming with fixed saliences for particular features is to prove that they are able to track the task—that is, that they continue to master the classification problem despite removing those cues that originally facilitated learning. Indeed, deprived of the intensity cue that was important for the successful classification, as demonstrated in a test including ambiguous images in this respect, pigeons were able to relearn the task (Troje et al., 1999). It would be interesting to determine not only what features they now used in an intensity-normalized situation but, particularly, whether those features were informative in the sense that they accurately divided the feature space into the experimenter-defined categories.

Given the complexity of the information that is contained in the surface properties of the face images, it is first necessary to reduce the dimensionality of the given picture set in order to find those dimensions that account for most of the variance in the stimulus space. *Principal component analysis* (PCA)—a multivariate statistical method first described by Pearson (1901) and thereafter widely used for such purposes (Sirovich & Kirby, 1987)—is a tool that yields an orthogonal basis with the axes ordered according to their overall variance. PCA has also been applied to similar tasks, like human face recognition (Turk & Pentland, 1991) and gender classification (O'Toole, Abdi, Deffenbacher, & Bartlett, 1991). Here, we applied PCA to our stimulus images in order to find those dimensions of variance that account for most of the variance *within* the complete picture set (males plus females).

A commonly used statistical method for finding the parameters in a multidimensional feature space that accounts for most of the variance *between* two or more groups of interest (here the two classes "Male" and "Female") is *discriminant function analysis*. We applied this method here in order to determine whether the dimensions that show the greatest contribution to the male/female separation are actually used as discrimination cues by the pigeons.

To investigate this possibility we used three related tests. First we examined rank correlations between response rates to individual stimuli and the values of the first 20 principal components for the same stimuli. Then we created synthetic test images that varied only according to each of the first three principal components and measured the spontaneous performance on those images. Furthermore, we assessed the role of colour by examining a correlation analysis with colour values of individual faces, and by then presenting the pigeons with greyscale test images. Finally, in order to determine whether any high-level features of the faces were used by the pigeons as discrimination cues, we "degraded" the images by using two standard filter techniques (blurring and block portraying) and measured the transfer ability to those filtered faces.

Method

Subjects

Experimental subjects were 11 pigeons (*Columba livia*) of a traditional livestock race ("Strasser") from a local supplier. These individuals had been used previously in Experiment 5 of Troje et al. (1999) and were selected out of a total of 15 subjects because they had reached a significant level of performance at the end of that experiment. They were housed in several compartments of a large (54 m²) outdoor aviary in groups of 8 to 10 individuals of mixed sex. Unlimited water and grit were provided daily in the aviary, whereas food was provided only in the learning boxes during and after the experimental sessions. The birds were maintained at weights slightly lower than their free-feeding weights.

Apparatus

The experiment was conducted in two identical one-key operant Skinner boxes located in the indoor laboratory, each box was connected through a passageway system with one outdoor aviary compartment (Huber, 1994). The front panel of such a 50 × 30 × 40 cm wooden chamber was

equipped with a pecking key and a food hopper. The oversized, clear Perspex pecking key (5-cm diameter, ENV-125M, MED Associates, U.S.A.) was mounted in the centre of the panel 28 cm above the floor. A 6×6 cm aperture for the food hopper was directly below the key (8.5 cm above the floor). The pigeon grain feeder (ENV-205M) used a 28-V DC solenoid-activated hopper designed to ensure reliable delivery of grain to the bird. A hopper light illuminated the receptacle area whenever grain was accessible. During the experiments, the chamber was weakly illuminated by a 2-W house-light (ENV-215) located in the rear part of the chamber. All images were presented at a size of 64×64 pixels (about 2×2 cm) in full colour (except in the greyscale test) against a black background on a 15" (38 cm, diagonal) PC monitor (Panasonic PanaSync 4G) at a distance of 5 cm behind the clear pecking key. The stimulus presentation procedure was controlled by Pentium PCs, one for each chamber, running a Delphi program that selected the patterns according to a pre-specified sequence. The MED behavioural control package, including a computer interface card (DIG-704) inside a Pentium PC, a modular interface (SG series, DIG-711, DIG-721), and the behavioural control software (MED-PC for Windows), was used to control events in the operant chambers and to register responses.

Stimuli

For the generation of the face stimuli, we used a database of computer models, which had been collected from three-dimensional laser scans of the heads of 100 male and 100 female students. The students' faces were free of any kind of accessories like glasses or earrings. Men were shaved and the hair of the head was digitally removed from the models (for details, see Troje & Bühlhoff, 1996). The 200 face models were randomly subdivided into two groups (Set A and Set B), each containing 50 male and 50 female faces. For each pigeon, one set served as the training set and the other as the test set.

The 3D faces served as models to render photorealistic images in frontal view using only ambient light and a black background (Figure 1a. If you are looking at this paper online, please click here to see a colour version of Figure 1). Those "original" images served only as a starting point for the generation of further stimulus sets used in the present experiments. The first set was derived by normalizing the images both with respect to their shape and with respect to their average intensity. Examples for such intensity-normalized, texture-only images (subsequently called "normalized") are shown in Figure 1b.

The shape normalization procedure is described in detail in Vetter and Troje (1995). It basically involves a non-linear deformation (i.e., a warp) of each single image. The deformation is defined such that the warped image would match with a predefined "prototype" image on a feature-by-feature basis. The prototype could, in principal, be any face. For computational reasons, however, it was chosen to represent the average of all the 200 faces in the database. The deformation fields were determined by using an adapted optical flow algorithm.

After the images had been normalized with respect to their shape, they were normalized with respect to their average intensity. This was done by first computing the average pixel value in each image and then rescaling the pixel values such that all images eventually had the same average pixel value, namely the overall average pixel intensity of the whole data set.

Due to the fact that the location of all features in the shape-normalized images are identical across the database, any linear combination of existing images will result in a new image that is qualitatively indistinguishable from the original images. Therefore, linear operations such as PCA are well defined on this image set. PCA is a linear decomposition that yields an orthogonal basis for a given data set with the origin centred at the mean of the data set and the axes (i.e., the "principal components") ordered with respect to their contribution to the overall variance. The first principal component is found by maximizing the variance of the data set as projected onto this axis. Each subsequent

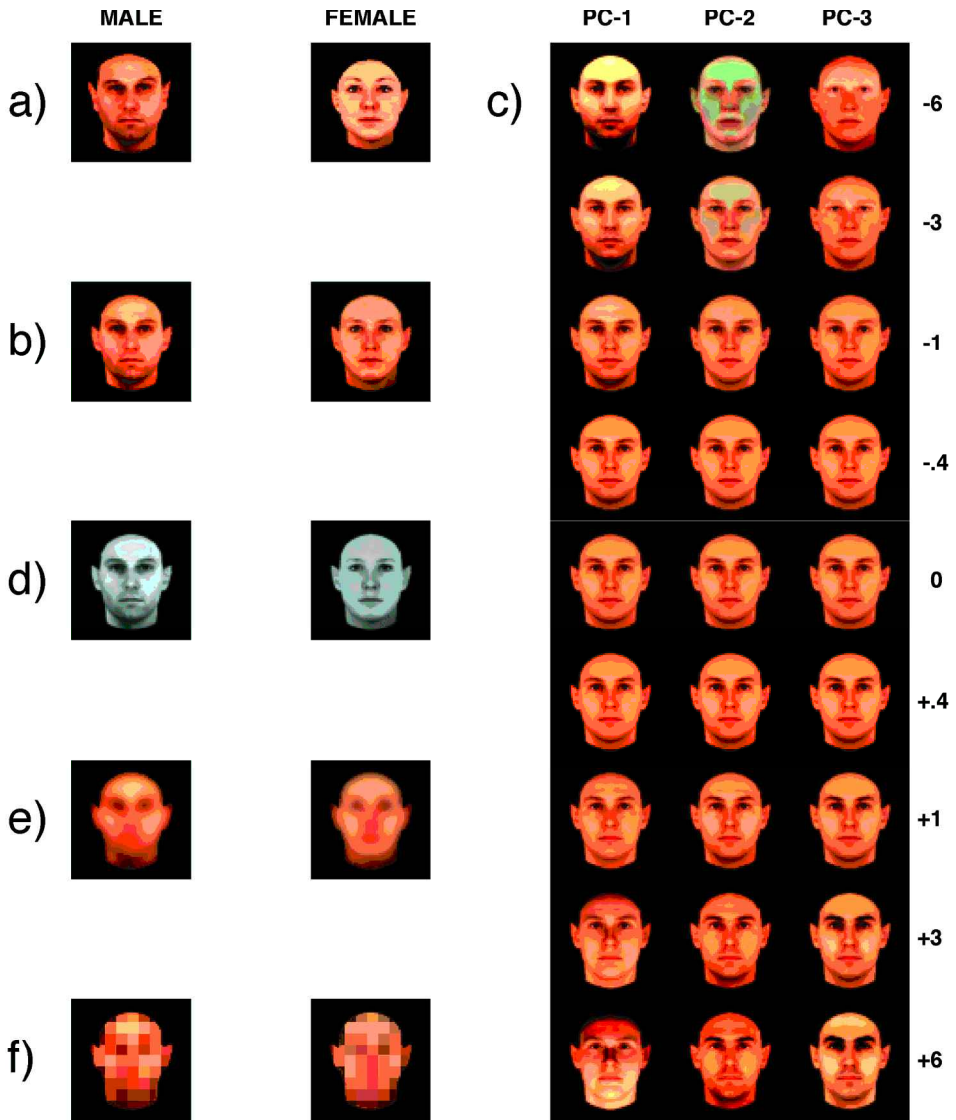


Figure 1. Examples of the face images. One example of a male (left) and a female (right) face is presented in the “original” representation (1a), in the “normalized” version (differences of shape and intensity are removed, 1b), in the greyscale version (1d), overlaid with a Gaussian filter (kernel radius = 20 pixel, 1e), and overlaid with a mosaic filter (block size = 6×6 pixel, 1f). On the right half of the figure (1c), examples of test images from the first three principal components are shown. To the average face (0) we added the respective normalized principal component with weights corresponding to 0.4, 1.0, 3.0, and 6.0 standard deviations away in both directions.

principal component is orthogonal to all previous axes and—given that constraint—again maximizes the variance of the data projected onto it. We computed a PCA using the raw image data as an input. Each image was coded by simply concatenating the three colour values for each pixel, which results in a vector with $64 \times 64 \times 3 = 12,288$ dimensions. The resulting principal components were then used

to generate images corresponding to points that were 0.2, 0.4, 0.6, 0.8, 1.0, 2.0, 3.0, 4.0, 5.0, and 6.0 standard deviations away from the average face (the origin, see earlier) in both directions along the first three components.

Figure 1c shows examples of such stimuli. In addition to the average face this figure depicts the synthetic images with weights corresponding to 0.4, 1.0, 3.0, and 6.0 standard deviations both above and below the mean along the first three principal components. The exaggerated faces at the extreme positions above and below the mean are especially informative; from these faces we gain some understanding of the information contained in the different principal components.

The first component picks up on a relative intensity difference between the upper and the lower part of the face. The second component picks up colour differences that seem to correlate with perceived gender but also on differences in the contrast of facial features. The higher principal components represent feature compositions that are harder to characterize.

The normalized images were also used to derive a set of greyscale images and a set of images with reduced spatial resolution. The greyscale images (Figure 1d) were derived by substituting each RGB colour value with a single grey pixel whose value (I) was given as a linear combination of the three colour values: $I = 0.3 * R + 0.59 * G + 0.11 * B$. Because we used only 100 from a total of 200 faces during the classification training, we were able to use the remaining 100 as novel faces for the greyscale manipulation.

Finally, we generated a further set of images with reduced spatial resolution by applying a set of Gaussian filters and a set of mosaic filters to the normalized images. The images were derived from a set of training faces that contained five male and five female faces. For each subject we used different sets, namely those faces that were discriminated with the highest accuracy (most pecks in the case of positives and fewest pecks in the case of negatives) during the last 16 presentations of classification training. Ten different filters were used. Five were mosaic filters that resulted in images with areas as large as 2×2 , 3×3 , 4×4 , 5×5 , or 6×6 pixels being replaced by the average intensity and colour of this region. The other five filters were Gaussian filters that would replace each single pixel by the weighted average of its neighbourhood. The weighting was defined by means of a Gaussian function with standard deviation of 4, 8, 12, 16, or 20 pixels. Applying those 10 filters to the 10 images we were able to show a total of 100 filtered faces to each subject. All colour images were shown in a resolution of 64×64 pixels and with 24 bits colour depth. The greyscale images had only 8 bits per pixel.

Procedure

The experiment required a discrimination between various face images. Before being exposed to an experimental problem, pigeons were trained to enter the box voluntarily (a description of this voluntary training method can be found in Huber, 1994) to find food in the hopper, and finally they were autoshaped by standard methods to peck at the illuminated key. We used a variant of the successive go/no-go procedure (described in Vaughan & Greene, 1984) as our discrimination method, which required the birds to peck in the presence of positive stimuli and to withhold pecking in the presence of negative stimuli. The standard training or test unit is a daily session consisting of 40 trials.

A training trial started with the onset of the stimulus and finished with the fulfilment of the response requirement. During this time the houselight was off. Pecks were counted only in the first 10 s of a trial. The images remained visible at least during a further variable interval (mean: 10 s, range: 1–20 s) and then until the response requirement was accomplished. If the stimulus was positive, the first response to occur within 2 s of a previous response terminated the stimulus presentation and produced 5 s of access to food. In contrast, responding to a negative pattern caused a continuation of the stimulus presentation, which terminated only if 8 s had passed without a

response occurring. In such trials no reinforcers at all were delivered. Test trials ran with “neutral” contingency—that is, they were terminated immediately after the first 10 s, during which pecks were counted, independent of whether the subject responded or not. Each trial was followed by a 4-s intertrial interval, during which the houselight was turned on. The intertrial interval ended with a short dark phase that signalled the presentation of the next stimulus.

Training sessions consisted of 20 positive and 20 negative trials in quasi-random order. In fact, we prespecified ten 40-trial sequences that were equalized with respect to either win/shift or win/stay strategies (Levine, 1959). Sessions conducted in the test phases contained both training and test trials. The test trials were randomly intermixed between the training trials, with the constraint that the first three trials were training trials and no more than two test trials could occur in succession.

The data reported here are from three tests conducted after training the pigeons to discriminate between 50 male and 50 female normalized images of human faces (Troje et al., 1999). Those subjects were arbitrarily assigned to the two different stimulus sets (Set A and Set B) and the two different reinforcement conditions (male-positive/female-negative and vice versa). The training consisted of two phases. In the first, lasting for 40 sessions, a subset of 10 faces was presented four times in each 40-trial session. On consecutive days, different images were shown so that a pigeon completed a whole cycle of the 100 images from either Set A or Set B within 10 days. We ran four of these cycles. In the second training phase, the same 100 pictures were used, but now 40 different images were shown only once per session. This phase lasted for at least 10 sessions and was then continued until the subject performed 5 consecutive sessions in which significant discrimination was demonstrated ($p < .05$, Mann–Whitney U -Test). If this criterion was not fulfilled within 30 sessions, training was terminated.

The first test phase began immediately after the completion of discrimination training. This test involved 18 sessions, each containing 28 training trials (14 positives and 14 negatives) and 12 test trials (neutral contingency). In the test trials, synthetic faces that were created to vary along the first three principal components were shown. During each session we inserted test images from only one principal component. In total, 21 test faces from each component were used: the average face and 10 faces on each side of the mean (see earlier). Faces -1 , 0 , and 1 were shown six times during the test, the remaining faces only three times (in randomized order).

The second test phase required the subjects to respond spontaneously to greyscale face images. A total of 10 sessions with 30 training and 10 test trials was sufficient to show all 100 test images once. The third test phase—in which the images overlaid with the filter masks were shown—also involved 10 sessions of 30 training and 10 test trials each.

Data analysis

To evaluate the discrimination performance, we used the ρ (rho) discrimination index introduced by Herrnstein, Loveland, and Cable (1976). This value equals the probability that the rank for the response to a positive stimulus is above that to a negative. The ρ -value is derived from the Mann–Whitney U -statistic by dividing the value of U by the product of the numbers of positive and negative stimuli contributing to its calculation.

The pecking rate in response to each stimulus was automatically registered during the first 10 s of each stimulus. For each pigeon, each pecking rate was then divided by the average over all pecking rates measured from that bird during the current 40-trial session. All further calculations were based on these normalized pecking rates. Group tendencies were given as means \pm standard deviations. Analyses of variance (ANOVA) were used to compare the mean pecking rates. We compared the normalized pecking rates to positive and negative stimuli and to training and test stimuli. Because training stimuli were shown repeatedly during a test, but test stimuli were shown only once, we used

only the normalized pecking rate of the first presentation of any training stimulus in each test session in order to compare the responding to training and test stimuli.

Results

Classification training. Pigeons in this experiment successfully classified the normalized faces according to sex. The learning curves of the subjects show a negatively accelerated acquisition function typical for many natural classification tasks (Watanabe et al., 1993). We computed ρ -values based on the normalized pecking rate emitted during the last presentation of all 100 training stimuli (mean \pm SD for positives: 1.76 ± 0.46 ; negatives: 0.71 ± 0.24) to evaluate the performance at the end of classification training. Although individual performances varied considerably (ρ -values between 0.65 and 0.88), all subjects acquired a significant discrimination level ($p < .01$, Mann–Whitney U -test). The mean discrimination index, averaged across all 11 subjects, equals 0.80 ($p < .001$).

Principal component test. Training pigeons to discriminate between male and female faces that contain all the surface properties resulted in a substantial preference for the average intensity of a face to be used as discrimination cue (Troje et al., 1999; Experiment 4). If this cue is removed, which stimulus parameters of the surface domain, if any, would the pigeons use to successfully master the classification task?

Our approach to specifying the stimulus attributes that might have been used by the pigeons started with a PCA (based on image vectors, not on pigeon judgements). Then, in order to find out whether the pigeons specifically used certain principal components, we computed the rank order correlation between pecking rates for individual faces and their coordinates on the respective principal component. The pecking rate associated with each single image used for this calculation was the mean of the normalized pecking rates emitted to the last 16 presentations of each stimulus in the classification training.

To exclude the partial correlation between pecking rate and sex that is not due to the parameter under investigation, we computed the correlation separately for only the male and for only the female faces (Table 1). Pecking rates of almost all animals showed significant correlations with the second principal component and some also with the first and the third components. Correlations were not distributed evenly across the two stimulus classes; for all animals very high correlations with the second component were found in the case of female faces, but only for three subjects in the case of male faces. The same tendency was evident when we analysed the third component. A relatively equal distribution of correlations could be found only in the investigation of the first component.

As indicated by the inspection of the first few principal components, the pigeons thus seem to use colour, the relative difference between the intensity of the upper and lower face, and perhaps shading as cues to discriminate between male and female faces, in this order of preference. Additional support that colour was actually involved in the coding scheme of the pigeons' analyses of the face images could be obtained by computing the rank order correlations of the pecking rates of individual faces with the red, green, and blue values of the images. As with the second principal component, the correlation

TABLE 1
Spearman rank correlations between normalized pecking rates and the values of the first three principal components for individual pigeons

Subjects	Male faces			Female faces		
	PG-1	PG-2	PG-3	PG-1	PG-2	PG-3
T50	.22	-.76	.19	.30	-.73	.34
T68	.45	-.35	-.02	.45	-.53	.04
T69	.48	-.59	.04	.38	-.41	.41
T2	-.31	-.20	-.22	.21	.65	-.15
T3	.07	-.32	.05	-.20	-.72	.33
T52	-.29	.09	-.10	.19	.63	-.18
T54	-.12	-.35	.22	-.33	-.78	.35
T60	-.15	.23	-.17	.18	.72	-.30
T63	-.20	.18	.07	.23	.72	-.22
T67	-.52	.12	.04	.08	.41	-.19
T65	.28	-.26	.40	-.02	-.50	.26
Mean	.28	.31	.14	.23	.62	.25

Note: The correlation was computed separately for male ($n = 50$) and female ($n = 50$) faces. Correlation coefficients larger than .24, .33, and .45 correspond to significance levels of $p < .05$, $p < .01$, and $p < .001$, respectively. Correlation coefficients larger than .33 ($p < .01$) are given in bold face.

coefficients for the three colour components indicate that the colour information contained in female faces, much more than in male faces, accounts for substantial proportion of variance in the pigeons' pecking behaviour.

We sought to support our hypothesis that pigeons attend to the information described by the first three principal components by confronting the subjects with test images that represent defined points on single principal components. Figure 2 shows the mean standardized response rates of all 11 subjects as a function of distance from the average along the first three principal components in both directions. The response tendency to faces that were some distance from the average represents a ceiling and a floor effect. In contrast, a large proportion of variance of the pecking frequency in the presence of test images within the range of -1 to $+1$ standard deviations is accounted for by its linear relationship with the first and second principal components. Subjecting the peck rates to the 11 test faces in the range between -1 and $+1$ standard deviations to a regression analysis for the three components separately revealed high R^2 -values for the first two components (first: $R^2 = 78.6\%$, $F = 33.1$, $p < .001$; second: $R^2 = 93.6\%$, $F = 131$, $p < .0001$) but a low R^2 to the third component ($R^2 = 3.3\%$, $F = .31$, $p > .05$).

In order to determine how well the test images along the first three principal components have been discriminated, we computed the discrimination index (ρ) by ranking the response rates to the 12 test images corresponding to points that are 1, 2, 3, 4, 5, and 6 standard deviations away from the mean in both directions (see Figure 2, white squares). Because the test images were repeatedly shown during the test, we only used the response

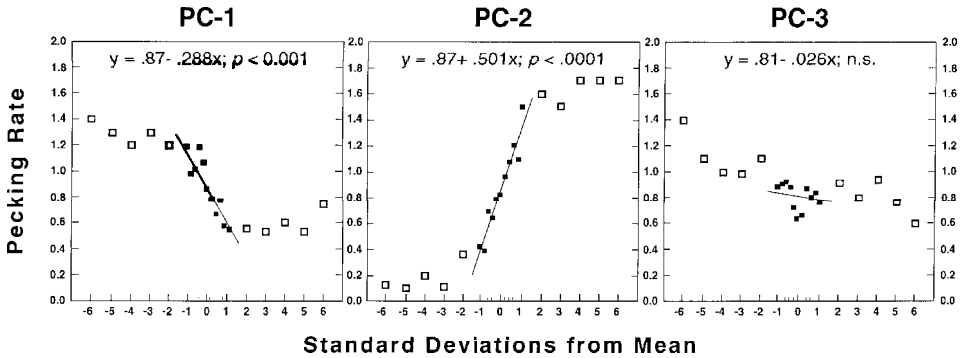


Figure 2. Normalized response rates as a function of distance from the average face (measured in standard deviations) along the first three principal components in both directions. We included in each graph the response rates (mean of 11 subjects) to synthetic faces that vary within a naturally occurring range (from -1 to $+1$) as well as those to the “exaggerated” faces that widely exceed that range. Note that the scale differs between these two subgroups. The regression lines are computed only for the first group of faces.

rates emitted during their first presentation. All 11 subjects were able to distinguish perfectly the test faces on either side of the mean of the second component; 10 pigeons achieved a ρ -value equalling 1.00 and one pigeon a ρ -value equalling 0.972 (mean ρ -value = 0.997). In comparison, the discrimination index for the training stimuli varied among individuals between 0.64 and 0.88 (mean ρ -value = 0.800). The mean discrimination index for the test stimuli of the first and third components were 0.894 and 0.745, respectively. Thus they responded to the variations in the first and second components as if they were exceptionally positive or negative. This was true for the second component even if we computed the discrimination index for all 20 test faces (mean ρ -value = 0.836). The respective ρ -values for the first and third components were 0.756 and 0.624, respectively.

Discriminant function analysis. Finding evidence that the pigeons primarily used the first two principal components is not necessarily indicative of their ability to track the task—that is, to use those feature dimensions that contribute to the separation of the sexes in the multidimensional face space. We therefore addressed the problem of how well males and females could be separated using a discriminant function analysis, given the measurements for the individual faces on several principal components.

We computed a discriminant function analysis for both image sets, Set A and Set B, each of which comprised 100 faces. In general, the main goal of such an analysis is to determine functions of the class descriptors (the principal components in this case) that maximally separate the groups. The simplest approach involves taking a linear combination of the variables with coefficients that are chosen to maximize the F -ratio for a one-way ANOVA for the variation within and between the groups. This *canonical discriminant function* was calculated using the first 10 principal components.

For both picture sets (A and B), the analyses yielded clear results. In each case, the means of the discriminant function differed significantly between groups: Set A, $\chi^2(10) = 85.26, p < .0001$; Set B, $\chi^2(10) = 135.86, p < .0001$. Computing the percentage of grouped cases correctly classified by the model reveals a value of 89% for Set A and 97% for Set B.

Most important, however, is whether the principal components that correlated highest with the pecking rates of individual subjects (PC2 and PC1, Table 1) and that yielded the highest R^2 -values in the principal component test (also PC2 and PC1) are actually those that contribute most to the male/female separation. In fact, we found in both picture sets that the second principal component contributed to the greatest variance between groups: Set A, $F(1, 98) = 49.56, p < .0001$; Set B, $F(1, 98) = 39.1, p < .0001$. Whereas in Set A the first component was the second greatest contributor to variation, $F(1, 98) = 21.64, p < .0001$, in Set B the first component was only the fifth greatest contributor, although it still yielded a significant F -ratio, $F(1, 98) = 5.46, p < .05$.

Greyscale test. Difference in colour served as a major cue for the discrimination. In view of the previous results it would be interesting to examine the effects on the pigeons' discrimination ability if colour were completely removed from the images. We measured the spontaneous response to greyscale images from 100 faces that were not shown during classification training. We expected that the generalization performance should be affected to an extent dependent on the control that colour exerted on the pigeons' classification performance.

The results are compatible with that hypothesis. In Figure 3, normalized pecking rates to positive and negative training stimuli, as well as to positive and negative test stimuli, for the 11 subjects are shown. Whereas the classification performance in the presence of training stimuli remained intact, transfer performance decreased to levels no better than chance. We computed a repeated measures, two-way ANOVA with class (positive vs. negative) and stimulus (training vs. test) as factors. This analysis revealed significant effects for both variables: class, $F(1,10) = 222.6, p < .001$; stimulus, $F(1,10) = 23.8, p < .001$, and also for the interaction between them, $F(1,10) = 162.2, p < .001$. The interaction indicated that the high level of performance for the training stimuli was not replicated for the test stimuli.

Filter test. In the filter test, the effects of altering some information contained in the images by blurring and by creating block portraits were examined. For each subject the five male and five female faces that were most accurately classified in training were

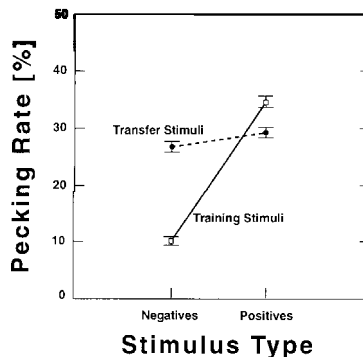


Figure 3. Normalized pecking rates emitted to positive and negative training stimuli and to positive and negative greyscale test stimuli in the greyscale test. Each data point contains the mean plus standard error of 11 subjects.

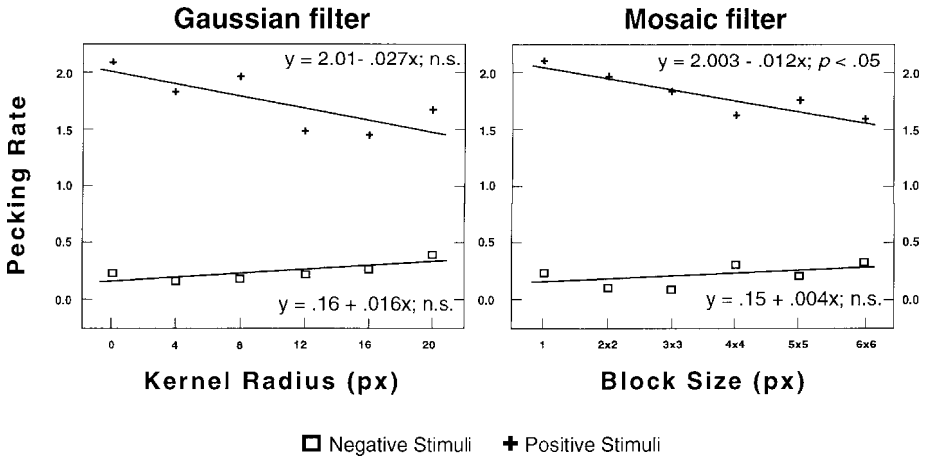


Figure 4. Mean normalized response rates as a function of filter radius (given in number of pixels) averaged over the five positive and five negative test stimuli, respectively.

transformed in five steps by increasing the diameter of the filter mask. Figure 4 shows the mean standardized response rates as a function of degree of manipulation averaged over the five positive and five negative test stimuli, respectively. The variance between individual subjects was low, and therefore we depict the means calculated across all subjects. It is evident from the regression lines fitted to the data that neither filter manipulation substantially affected the spontaneous classification ability of the pigeons. Only the decreasing frequency of pecking in the presence of positive test images that were presented as block portraits showed a significant trend, $R^2 = 71.6\%$, $F = 10.1$, $p < .05$. However, even in this case, the difference in pecking positives and negatives was highly significant, $F(1,10) = 93.5$, $p < .0001$.

GENERAL DISCUSSION

The central issue addressed here was to identify the information that pigeons use to learn a complex natural class discrimination. Although it has been known for more than 35 years that pigeons are able to categorize a wide variety of visual stimulus classes, two important issues arising from the core question of categorization have remained: (1) when and how differential reinforcement shapes the features entering into a polymorphous rule and (2) what the perceptual dimensions are and how they can be physically represented (Herrnstein, 1984).

Together with the findings of the preceding study (Troje et al., 1999) we are able to offer some answers to these questions. In the Troje et al. study we already proved the pigeons' capacity for open-ended categorization by showing very quick and accurate learning to sort 100 human faces with respect to their sex and by showing perfect transfer performance to 100 novel instances. However, only pigeons that were provided with the surface information of the pictures were successful at the discrimination task. Moreover, we found that accurate discrimination by those pigeons was primarily controlled by the

average intensity of the faces, which proved to be a valuable cue for the sex discrimination. This result is easily explained by feature analysis—that is, by suggesting that pigeons came with a fixed saliency for the intensity feature and therefore saw new instances of male and female faces as being similar to training stimuli unconditionally (Lea, 1984). We would thus have good answers to the earlier Question 2 but not to Question 1.

Therefore, the present experiment was aimed at determining not only what features the pigeons now used in an intensity-normalized situation but, particularly, whether those features were informative in the sense that they accurately divided the feature space into the experimenter-defined categories. Applying principal component analysis to the human face images, we reduced the dimensionality of the complex face space from many thousand possible lines of variation to a few variables that provided for most of the variation inherent in these images (irrespective of their sex). Then we employed discriminant function analysis to identify those principal components that provided for most of the variation between male and female faces. Calculating—separately for both classes—rank order correlations between response rates to individual faces and the values of three potent “class-separator” components for the same 100 stimuli revealed that the pigeons’ classification behaviour was actually controlled by those “category features”.

From inspection of synthetic faces that were created as extreme instances along the principal component dimensions we “described” the three most important dimensions as representing differences in *colour*, *intensity gradients*, and *local shading* between male and female faces. The very high proportion of variation in responding accounted by the three colour components when used for the correlation analysis instead of the principal component values supports our hypothesis concerning the predominant role of colour.

We also obtained direct support from presenting synthetic faces varying only along each of the first three principal components. Response rates to stimuli lying within the normally occurring range in training (between -1 and $+1$ standard deviation away from the mean) were linearly related (R^2 up to 93.6%) to their position on the components. Even more conclusive, when we presented the “exaggerated” stimuli that deviated considerably from the mean on each dimension (up to -6 and $+6$ standard deviations away from the mean, respectively), categorization performance was not disturbed but, to the contrary, was enhanced. Thus they responded to those “single feature stimuli” as if they were exceptionally positive or negative. In terms of the ethologists these test stimuli functioned like “supernormal” stimuli or “superreleasers” (Tinbergen, 1951; see also Lea & Harrison, 1978). We may also think of a prototype effect or—according to the Spencian theory of gradient interaction—of a peak shift effect. To some extent our results are compatible with these theories, but what is interesting about the data is that there is no evidence of a decline in responding to $S+$, or of a recovery to $S-$, as the principal component stimuli become more extreme (see Huber & Lenz, 1996, for similar effects).

Finally, if a “polymorphous” combination of colour, intensity, and shading information describes the categorization rules formed by the pigeons, the subjects’ classification ability should be impaired proportionally to the reduction of that particular information. In contrast, filtering of the faces that leaves the three components of the pigeons’ class description unaffected should not influence the subjects’ classification ability. We conducted only three such tests by presenting greyscale versions, blurred versions, and block

portraits of faces. As predicted, although the loss of colour information produced substantial performance decrements in all subjects, neither blurring nor blocking led to a significant disruption of classification ability.

In conclusion, these findings are entirely consistent with the *feature learning* hypothesis of categorization (Herrnstein et al., 1976; Huber & Lenz, 1993; Jitsumori, 1993; Lea & Harrison, 1978; Lea & Ryan, 1983; Lubow, 1974; Morgan, Fitch, Holman, & Lea, 1976; see also Huber, 1999, 2000, for an extended discussion). Particularly, the “supernormal” responding to the “supernormal” PC stimuli, which are far away from the variances presented by the training stimuli, supports the feature learning explanation. The categorization process can be described as a kind of perceptual carpentry (Herrnstein, 1984)—that is, from the available stimulus dimensions, differential reinforcement selects those that differentiate between positive and negative instances. The resulting category features are then likely to be woven into a polymorphous rule for categorizing, even if item-specific information had also previously been stored in the picture memory (the dual storage process hypothesis, Cook, Wright, & Kendrick, 1990).

This conclusion is clearly in contrast to a pure template-based model of classification transfer, which holds that each pixel of the current input is compared with each pixel of the remembered form (Donis, Heinemann, & Chase, 1994). It is also difficult to see how an exemplar model derived from the key notion that judgements in classification tasks are based on retrieval of exemplar information rather than on category level information (Medin & Schaffer, 1978) accounts for the present data. On the other hand, modern formulations of the exemplar account, like the configural theories by Pearce (1987) and Kruschke (1992), assume that the exemplar representation contains information about the features that were present on a particular trial and that during training there is generalization from one exemplar to another according to the degree of their similarity. These theories allow that pigeons pay more attention to some features than others and that the category-relevant features—that is, those that occur more frequently in one class (or type of reinforcement) than in another—will exert greater control over responding than will category-irrelevant features—that is, those that occur in both classes.

So far, the present experiment cannot be used as a test between feature and exemplar theories. Such tests are manageable with a few stimuli representing compounds of only a few all-or-none elements (like the $A^+ BC^+ ABC^0$ discrimination; see Pearce, 1994). However, it will be difficult (if possible at all) to discriminate between these two types of theory—or to think of any experiments that could without question be used to choose between the two theories—when it comes to the analysis of categorization. As demonstrated here, categorization experiments may involve hundreds of naturalistic stimuli, which vary on a vast amount of feature dimensions and which also involve idiosyncratic aspects (occurring in single instances).

Nevertheless, the present data may not only provide novel insights into the features that control the discrimination of complex visual stimuli by pigeons but may also represent further steps towards a full understanding of visual categorization as a means to enact the correspondence between visual and natural categories.

REFERENCES

- Bergen, J.A., Anandan, P., Hanna, K.J., & Hingorani, R. (1992). Hierarchical-model-based motion estimation. In G. Sandini (Ed.), *Second European Conference on Computer Vision* (pp. 237–252). Berlin: Springer.
- Cook, R.C., Wright, A.A., & Kendrick, D.F. (1990). Visual categorization by pigeons. In M.L. Commons, R.J. Herrnstein, S. Kosslyn, & D. Mumford (Eds.), *Quantitative analysis of behavior* (Vol. 8, pp. 187–214). Cambridge, MA: Ballinger.
- Donis, F.J., Heinenmann, E.G., & Chase, S. (1994). Context effects in visual pattern recognition in pigeons. *Perception & Psychophysics*, *55*, 676–688.
- Fersen, L. von, & Lea, S.E.G. (1990). Category discrimination by pigeons using five polymorphous features. *Journal of the Experimental Analysis of Behavior*, *54*, 69–84.
- Gibson, J.J. (1979). *The ecological approach to visual perception*. Boston: Houghton Mifflin.
- Herrnstein, R.J. (1984). Objects, categories, and discriminative stimuli. In H.L. Roitblat, T.G. Bever, & H.S. Terrace (Eds.), *Animal cognition* (pp. 233–261). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- 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. (1994). Amelioration of laboratory conditions for pigeons. *Animal Welfare*, *3*, 321–324.
- Huber, L. (1999). Generic perception: Open-ended categorization of natural classes. *Current Psychology of Cognition*, *18*, 845–888.
- Huber, L. (2000). Visual categorization in pigeons. In R.G. Cook (Ed.), *Avian visual cognition* [On-line]. Available: <http://www.pigeon.psy.tufts.edu/avc/>
- Huber, L., & Lenz, R. (1993). A test of the linear feature model of polymorphous concept discrimination with pigeons. *The Quarterly Journal of Experimental Psychology*, *46B*, 1–18.
- Huber, L., & Lenz, R. (1996). Categorization of prototypical stimulus classes by pigeons. *The Quarterly Journal of Experimental Psychology*, *49B*, 111–133.
- 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., & Yoshihara, M. (1997). Categorical discrimination of human facial expressions by pigeons: A test of the linear feature model. *The Quarterly Journal of Experimental Psychology*, *50B*, 253–268.
- Kruschke, J.K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review*, *99*, 22–44.
- 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. *The Quarterly Journal of Experimental Psychology*, *30*, 521–537.
- 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.), *Discrimination processes* (Vol. 4, pp. 263–276). Cambridge, MA: Ballinger.
- Levine, M. (1959). A model of hypothesis behavior in discrimination learning set. *Psychological Review*, *66*(6), 353–366.
- Lubow, R.E. (1974). High-order concept formation in the pigeon. *Journal of the Experimental Analysis of Behavior*, *21*, 475–483.
- Medin, D.L., & Schaffer, M.M. (1978). Context theory of classification learning. *Psychological Review*, *85*, 207–238.
- Morgan, M.J., Fitch, M.D., Holman, M.D., & Lea, S.E.G. (1976). Pigeons learn the concept of an "A". *Perception*, *5*, 57–66.
- O'Toole, A.J., Abdi, H., Deffenbacher, K.A., & Bartlett, J.C. (1991). Classifying faces by face and sex using an autoassociative memory trained for recognition. In K.J. Hammond & D. Gentner (Eds.), *Proceedings of the thirteenth annual conference of the Cognitive Science Society* (pp. 847–851). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Pearce, J.M. (1987). A model of stimulus generalization for Pavlovian conditioning. *Psychological Review*, *94*, 61–73.

- Pearce, J.M. (1994). Similarity and discrimination: A selective review and a connectionist model. *Psychological Review*, *101*, 587–607.
- Pearson, K. (1901). On lines and planes of closest fit to a system of points in space. *Philosophical Magazine*, *2*, 557–572.
- Sirovich, L., & Kirby, M. (1987). Low-dimensional procedure for the characterization of human faces. *Journal of the Optical Society of America A*, *4*, 519–554.
- Tinbergen, N. (1951). *The study of instinct*. Oxford: Clarendon.
- Troje, N., & Bühlhoff, H. (1996). Face recognition under varying pose: The role of texture and shape. *Vision Research*, *36*, 1761–1771.
- 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.
- Troje, N., & Vetter, T. (1998). Representations of human faces. In C. Taddei-Ferretti & C. Musio (Eds.), *Downward processing in the perception representation mechanism* (pp. 189–205). Singapore, New Jersey, London, Hong Kong: World Scientific.
- Turk, M., & Pentland, A. (1991). Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, *3*, 71–86.
- Vaughan, W.J., & Greene, S.L. (1984). Pigeon visual memory capacity. *Journal of Experimental Psychology: Animal Behavior Processes*, *10*, 256–271.
- Vetter, T., & Troje, N. (1995). Separation of texture and two-dimensional shape in images of human faces. In G. Sagerer, S. Posch, & F. Kummert (Eds.), *Mustererkennung 95* (pp. 118–125) Germany: Springer Verlag.
- Vetter, T., & Troje, N. (1997). Separation of texture and shape in images of faces for image coding and synthesis. *Journal of the Optical Society of America A*, *14*, 2152–2161.
- Wasserman, E.A., Tassinari, L.G., Bhatt, R.S., & Sayasenh, P. (1989). *Pigeons can discriminate emotional expression and individual identity from photographs of the human face*. Paper presented at the Annual meeting of the Psychonomic Society, Atlanta, GA.
- Watanabe, S., Lea, S.E.G., & Dittrich, W.H. (1993). What can we learn from experiments on pigeon concept discrimination? In H.P. Zeigler & H.-J. Bischof (Eds.), *Vision, brain, and behavior in birds* (pp. 351–376). Cambridge, MA: MIT Press.

Manuscript received March 10 2000
Accepted revision received June 16 2000