

XOR style tasks for testing visual object processing in monkeys [☆]

Britt Anderson ^{a,1}, Jessie J. Peissig ^{b,2}, Jedediah Singer ^{a,1}, David L. Sheinberg ^{b,*,2}

^a *Brain Science Program, Box 1953, Brown University, Providence RI 02912, USA*

^b *Department of Neuroscience, Box 1953, Brown University, Providence RI 02912, USA*

Received 10 June 2005; received in revised form 9 September 2005

Abstract

Using visually complex stimuli, three monkeys learned visual exclusive-or (XOR) tasks that required detecting two way visual feature conjunctions. Monkeys with passive exposure to the test images, or prior experience, were quicker to acquire an XOR style task. Training on each pairwise comparison of the stimuli to be used in an XOR task provided nearly complete transfer when stimuli became intermingled in the full XOR task. Task mastery took longer, accuracy was lower, and response times were slower for conjunction stimuli. Rotating features of the XOR stimuli did not adversely effect recognition speed or accuracy.

© 2005 Elsevier Ltd. All rights reserved.

Keywords: Conjunction; Monkey; Visual; Behavior; Learning

1. Introduction

“Binding by synchrony” is a current, but controversial, hypothesis for explaining how the brain represents visual objects (Shadlen and Movshon, 1999; Singer, 1999). Binding by synchrony proposes that the synchronous occurrence of neuronal action potentials is of particular importance in tagging which neurons go with which and consequently which visual features are bound together. Testing this theory is conceptually straightforward; one measures the joint firing statistics of multiple neurons when a subject is viewing the same visual features in two conditions: one where they do, and one where they do not “go together.” Operationalizing these ideas, however, produces substantial practical problems. The one which we focus on here is the need,

in the case of primate neurophysiology, for a monkey to solve a non-linear mapping from stimulus to response.

In a typical monkey, visual classification task, images of objects are displayed on a computer screen and the monkey is required to make a physical response by touching the screen, pressing a lever, or looking at a specific target. If the monkey is more accurate than chance, then the experimenter asserts that the image is being correctly recognized. When object identification is studied at the individual level, it is common to have several, but almost always many less than one hundred, images from the same basic category (e.g., Kobatake, Wang, & Tanaka, 1998; Sigala, 2004). When the objects in the set appear to us as visually similar, we may believe that the monkeys’ responses are based on more than individual diagnostic elements, but we do not actually know this. It is still possible for such relatively small sets of objects, there is, for each image, a unique feature that allows classification. What the monkeys may learn through our training is which features are diagnostic for which images. To evaluate the neurological mechanisms of visual feature binding, we must guarantee that this is not the case. There must be, at a minimum, at least two areas of the image necessary for correct categorization. However, this requirement introduces a new difficulty: if we want any

[☆] We thank Julie Lamin for assistance with animal care and testing and Ryan Mruczek and Keisuke Kawasawki for helpful comments and advice.

* Corresponding author. Tel.: +1 401 863 9575.

E-mail address: britt@brown.edu (B. Anderson).

¹ Supported by the Burroughs–Wellcome Fund.

² Partially supported by the James S. McDonnell Foundation, the Sloan Foundation, and NIH EY014681.

one feature to be uninformative, but have correct responses require the conjunction of two features, we are in the domain of non-linearly separable mappings. The canonical example of a non-linear mapping problem is the exclusive-or (XOR). In an XOR task, the combinations of either both (1 and 1) or neither (0 and 0) stimuli are mapped to one response and the either conditions (1 and 0 or 0 and 1) are mapped to the alternative response (see also Fig. 1). There is little data on explicit solutions of non-linearly separable problems in animals and the little data available suggests that they are hard, especially for monkeys.

Many studies on learning non-linear mappings use implicit learning techniques in the tradition of classical conditioning. Two such techniques are biconditional discrimination and negative patterning. Biconditional discrimination had been shown for rabbits (Saavedra, 1975), pigeons (Rescorla, Grau, & Durlach, 1985), monkeys (Saunders & Weiskrantz, 1989), and humans. For example, Lober and Lachnit (2002) successfully trained people to associate two-letter strings (B, G, T, and X) to either shock (reinforcement condition) or no shock (no reinforcement condition) and measured galvanic skin conductance changes. Negative patterning is even more clearly an XOR like task (Kehoe & Macrae, 2002). In a study of eye blink conditioning, Kehoe and Graham (1988) exposed rabbits to cross-modal combinations of tone and light stimuli. They observed responses to the stimuli individually and declining responses to the conjunction. Similar results using two auditory stimuli and reward with food have been shown in rats, pigeons, and bees (Deisig, Lachnit, Giurfa, & Hellstern, 2001; Redhead & Pearce, 1995).

Studies showing explicit learning of an XOR like task in primates are few. Thorpe, O'Regan, and Pouget (1989) displayed a 4×4 grid of lights that flashed patterns at 5/s; all patterns were evaluated according to a specific rule, one of which followed an XOR pattern. The human subject was never able to learn the XOR mappings. Baker, Behrmann, and Olson (2002) used an XOR task as a component of an electrophysiological study of monkey inferotemporal cortex. Stimuli were simple geometric patterns connected by a central baton (similar to those used in Experiment 3 of this report). Learning for one monkey took ~ 5000 trials per stimulus for a set of eight objects and the other monkey required ~ 7000 for another set of eight objects. Compared to conventional classification tasks, this is a large number of trials. This result is in accord with those of Smith, Minda, and Washburn (2004). They assessed the ability of four monkeys to learn a variety of problems using simple geometric shapes of varying size and color. One of the problems was of the XOR type and was more difficult for monkeys, relative to other learning tasks, than for a comparison group of human subjects. The *unitization* tasks used by Goldstone (2000) to test normal human subjects are also similar to XOR tasks in that the “squiggles” used required that multiple pieces be recognized for correct classification. Performance was shown to vary with the number of conjunctions required, but Goldstone argued that eventually response times became equivalent across stimuli suggesting that functionally a unitary representation had been performed.

The common result of these studies is that XOR tasks are hard for primates (people and monkeys), often

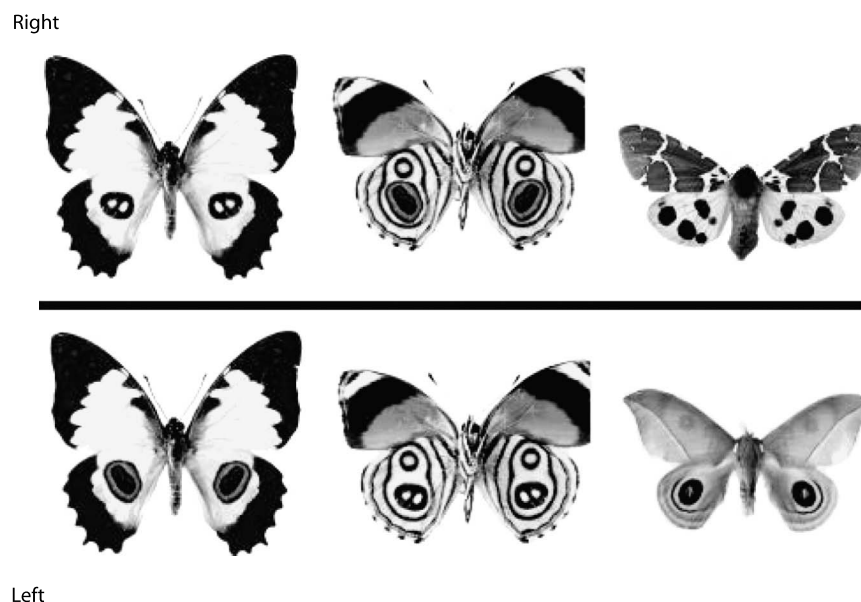


Fig. 1. The six stimuli used in Experiment 1. The top row shows the stimuli associated with the right button press and the bottom row those associated with the left button press. The 2 moths on the far right were unique stimuli and provided an index for how quickly traditional stimuli were learned in comparison to the XOR stimuli. The four butterflies on the left are the XOR set. The features are the butterfly body: white or orange; and the tail spots: brown or black with white spots. What makes the task an XOR task is that neither the body type or tail spot type alone allow responding above chance; the combination is needed to determine the correct button press. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this paper.)

requiring thousands of trials to master. All of these studies used simple geometric figures, lines, or shapes. Could a richer visual stimulus improve performance, making these types of tasks more practical? Or could prior experience with the objects to be used in an XOR task support more efficient performance. To reiterate, XOR tasks, in principal, offer an excellent task type for exploring the neural correlates of “binding,” but are they practical? While demonstrating that monkeys can master such tasks, prior research has suggested very long learning times and have used a relatively impoverished set of visual stimuli. In the present report, we look at several manipulations of the basic XOR paradigm to enhance their practicality. In addition, since an implicit assumption of many investigations of object classification is that stimuli are functionally “unitized,” we examined if frequent prolonged training with objects defined by configurations of object parts would be treated, by the monkeys, similarly to objects where classification could proceed by diagnostic features.

2. Experiment 1

Can monkeys learn an XOR task, where the test stimuli are complex pictures? One might think that complicated images could make finding the relevant pieces more difficult and solving the problem harder. On the other hand, more natural appearing images might be easier for monkeys than abstract simple shapes. We tested three monkeys for their ability to learn to press left or right buttons to specific pictures of butterflies and moths; some of the pictures were hybrids (Fig. 1) that followed an XOR mapping. We used butterflies as stimuli because they are naturally occurring objects, have many potentially salient features, and are a source of food for some species of macaque (Kumara, Singh, Sharma, Singh, & Kumar, 2000). In addition, these three monkeys had considerable experience with learning to classify similar images of butterflies in prior experiments and were highly proficient.

2.1. Methods

2.1.1. General

These experiments used three adult male rhesus macaques (wt: 9–13 kg; identifiers J, S, and T) housed and trained in accordance with the policies and procedures set forth in the US Public Health Service Policy on Humane Care and Use of Laboratory Animals and the National Institutes of Health Guide for the Care and Use of Laboratory Animals, as adopted by the Society for Neuroscience in its Policy on the Use of Animals in Neuroscience Research. All experiments were approved by the Institutional Animal Care and Use Committee.

All monkeys had titanium head posts which were used to maintain head restraint during behavioral testing. The combined duration of daily test sessions was between 1 and 3 h. Electrically shielded and sound isolated testing rooms were equipped with graphics accelerated computers for image

display. The monitoring room contained the control computers for monitoring event timing. Both a Robinson style eye coil systems (CNC Engineering; used for J), and an ISCAN RK-726PCI video eye tracking system (ISCAN, Burlington, MA; used for S and T) were employed.

All trials, except as specifically mentioned below, began with a warning tone and required acquisition of a central fixation spot followed by maintained fixation for 450 ms. After a delay of 200 ms, the trial stimuli were displayed on a computer screen. The monkeys responded by pressing one of two buttons. Correct trials were rewarded with fruit juice (or water) and a low pitched tone. Incorrect trials resulted in no juice and a higher pitched tone. The intertrial intervals were 1000 ms.

Stimuli images were selected from a photo database (Hemera Photo Objects Premium Image Collections I and II; Hemera Technologies Corporation, Seattle, Washington; Experiments 1 and 4), produced with 3DStudioMax (Discreet, Montreal, Canada; Experiment 2) or rendered using our local OpenGL based stimulus presentation software (Experiment 3). Some of the images were manipulated in Adobe Photoshop (Adobe, San Jose, CA). The hardware for stimulus presentation was a dedicated graphics workstation at a resolution of 1280 × 1024 and a screen refresh rate of 100 Hz. For behavioral control of the experiments, a network of interconnected PCs running the QNX real time OS (QSSL, Ontario, Canada) was used.

2.1.2. Specific

For the XOR classification task, each of six images was associated with either a right or left button press (Fig. 1). Two of the images were independent from all the others



Fig. 2. One example of the ring removed stimuli. In principle, the area where the tail spot abutted the butterfly body could provide a possible diagnostic location, that is a small single area of the image would be unique for each stimulus. To test whether monkeys J or S used this approach to stimulus recognition we tested their performance on images with and without rings removed.

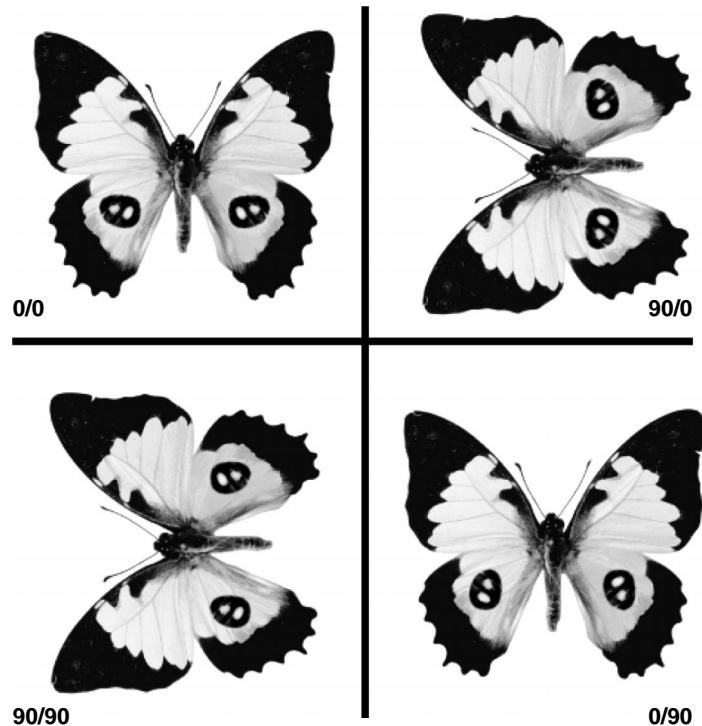


Fig. 3. An example of the rotated versions of the stimuli used in Experiment 1. We prepared stimuli in which the tail spot was rotated relative to the main axis of the body by 15 or 90° (bottom). We then presented these images in the original orientation (lower right) or rotated in the opposite direction 15 or 90° so that the tail spot could be in its original alignment but the body would not be (lower left). Since no single rotation could bring the images with the tail spots rotated relative to the body back into the learned alignment, the effect of these rotations on accuracy and speed of recognition represented a test of the configural hypothesis.

and provided an index of how quickly a simple diagnostic version of these stimuli could be learned. The other four images were composites in which tail spots, cut from two other butterfly pictures, were blended onto the bodies of two other butterflies. The button mapping associated with these four pictures created an XOR task in which it was necessary to identify the conjunction of the tail spot and some element of the butterfly body to be able to correctly classify the image.

For two of the monkeys (S and J) additional conditions were administered. To explore whether the monkeys might be responding to a limited area of the unique regions where the tail spot abutted the surrounding body six additional versions of the images were made in which a small ring was cut out of the pictures of the butterflies surrounding the tail spots. The response mapping remained the same (Fig. 2).

We also wished to test whether the images of butterflies were being treated as integrated wholes or as separate parts. We developed sets of stimuli for each butterfly in which the tail spots were rotated either 15 or 90° relative to the main axis of the butterfly body (Fig. 3). We then presented the learned versions and the versions with the rotated tail spots upright and with the entire composite rotated 15–90°. If the images were being processed as wholes (body and tail spot as one) then the conditions where the tail spot was rotated relative to the body should be the slowest. If the two parts were being treated separately, then condition 90/0 should be the slowest since both the tail spot and the body would

have to undergo a rotation back to the canonical learned position before matching (Fig. 3).

2.2. Results and discussion

All three monkeys successfully learned the visual XOR task. The monkeys were slower to learn the conjunction stimuli (Fig. 4).³ This difference cannot be explained by the fact that there were four XOR stimuli and two unique stimuli as we have previously found that these monkeys can quickly learn larger sets of similar items (Peissig & Sheinberg, unpublished data). Monkey J appeared to learn the task more gradually than monkeys S and T, but all required several hundred repetitions to learn the conjunctive stimuli when they could learn the unique stimuli with just a few repetitions (e.g., monkey S learned the unique stimuli with less than 10 exposures). This, however, compares favorably with the thousands of trials administered in earlier studies

³ The experiments reported used image sets of different sizes and had images presented differing numbers of times. To report the results in a common format for all experiments, we present our graphical results as a moving average of “blocks.” Each block contains fifteen trials for each image in the particular group being plotted. These blocks used for the graphical presentation allow us to put data groups with unequal image set sizes on a common abscissa and to match groups for the number of exposures per image. These analysis blocks do not equate to calendar day or experimental blocks of trial. Where this distinction is important, we explicitly refer to experimental blocks, numbers of trials, or training days.

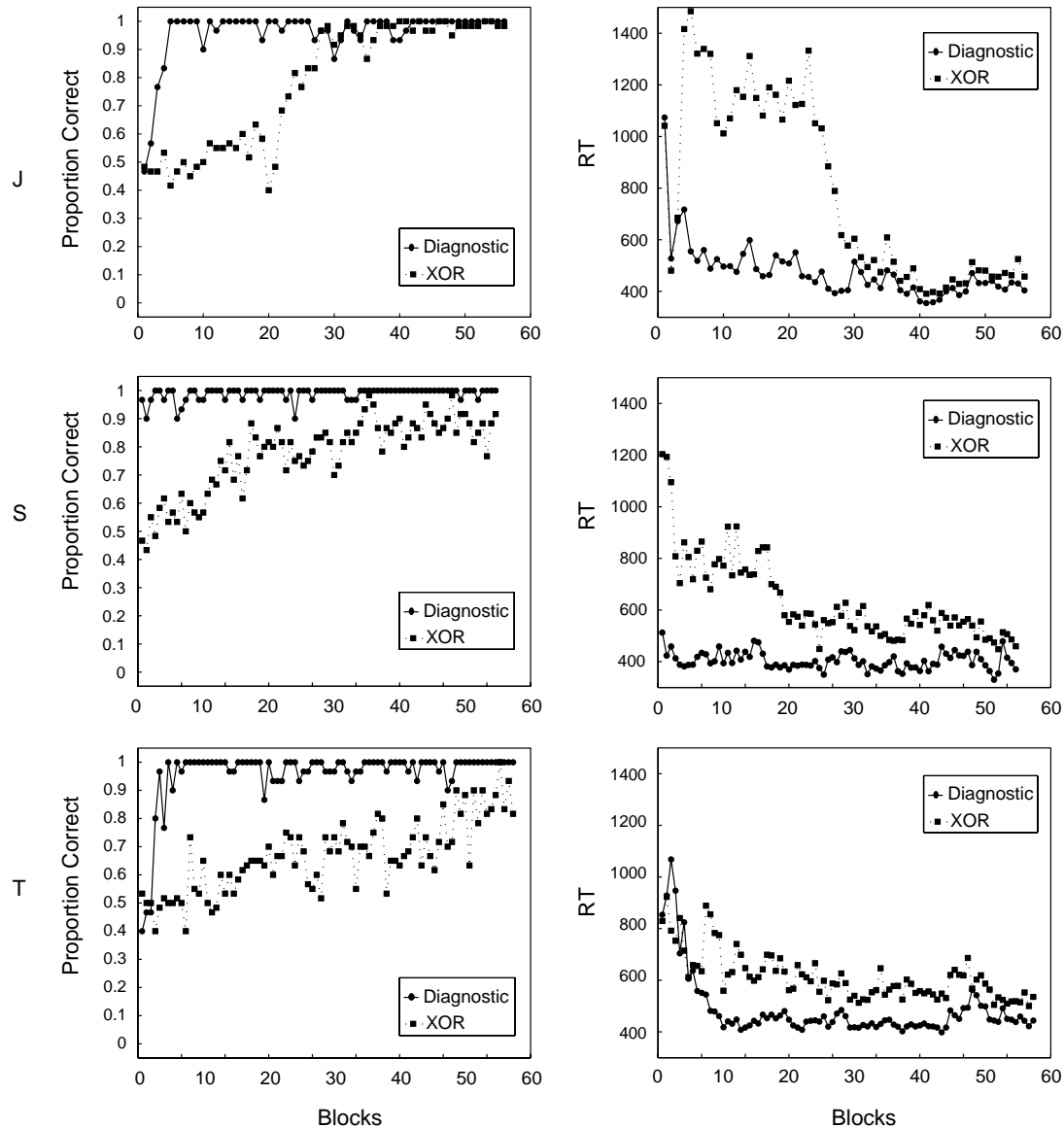


Fig. 4. Accuracy and speed of learning an XOR mapping. The results of each monkey is in a row (J, S, and T). Accuracy graphs are in the left column and response time graphs are in the right column. Each data point represents the mean (accuracy) or median (response time) of 15 images of each type in the particular class. This allows us to display the graphs for diagnostic and XOR stimuli on a common abscissa and to compare performance for the XOR and diagnostic groups for a similar number of exposures per image. Since there were four XOR stimuli each XOR data point represents 60 trials. Chance performance is 0.5.

(Baker, Olson, & Behrmann, 2004; Smith et al., 2004). Additionally, the conjunctive stimuli were not recognized as quickly or as accurately as the diagnostic stimuli. Focusing on the last 200 trials when each monkey was achieving a greater than 85% global accuracy, their response time to conjunction stimuli was slower than the diagnostic stimuli (permutation test, $p < 0.05$). During this period, they also committed more errors on conjunction stimuli. J committed only one error (a conjunction stimulus), S committed 0 errors for 66 diagnostic stimuli and 10 errors on 134 conjunction stimuli. For T the fractions were 1/68 diagnostic stimuli and 23/132 conjunction stimuli.

Despite our intentions, did the monkeys build “diagnostic” features for the XOR stimuli by focusing in on the

region where the tail spots and body abutted? In this region of the image, a relatively small zone could provide diagnostic evidence of the appropriate response.

For monkey J, there was no evidence of this strategy, but it appeared monkey S might be using such a strategy. For monkey J testing on stimuli with rings of the image around the tail spots removed had no effect on performance, but for S there was a significant decline in response accuracy (permutation test, $p = 0.001$; see Table 1 bottom row) suggesting that the two monkeys may have adopted different strategies. The effect for monkey S could not be attributed to a non-specific effect of manipulating the image since the decrement in accuracy was only noticed for the XOR stimuli (see Table 2). At the very least, these data confirm

Table 1
Performance for the ring condition overall

	J		S	
	RT	Accuracy	RT	Accuracy
No ring	444	0.995	513	0.96
Ring	458	0.963	569.5	0.734
<i>p</i> value	0.12	0.12	0.09	0.001

Table 2
Performance for the ring condition: accuracy S

	Class	
	Diagnostic	XOR
Ring absent	1.0	0.94
Ring present	1.0	0.61

heterogeneity at the level of individual monkeys in their strategies for solving visual object recognition tasks Keating and Keating, 1993.

Even after hundreds of exposures to each image, the response times for the unique and XOR images did not merge (Fig. 4). This raised the question of whether objects requiring feature conjunctions were being treated as single, composite, “things.” Motivated by the paradigmatic example of the face inversion effect for configural processing, we tested for evidence of holistic processing by using stimuli with rotated components. If the stimuli were being processed as wholes, rather than pieces, then rotations that broke the relationship between the body and the tail spots (e.g., conditions 90/90 and 0/90 in Fig. 3) should lead to slower, and possibly less accurate recognition. This would lead to an interaction effect for the conditions of body and tail spot rotation in the XOR stimuli class. We assessed monkeys J and S on rotated versions of the images. For fifteen degree rotations, there was neither an effect of whole object or part rotation nor any interaction [permutation ANOVA test; Pesarin (2001)]. When rotating parts of the butterflies a larger amount (90°), we found for XOR stimuli a main effect of rotation of the body, but no effect of rotating the tail spot alone nor an interaction effect (see Table 3 for statistical results). One difficulty with interpreting this result as a deficiency in holistic processing is that while the inversion effect can be shown for non-face stimuli (e.g., Farah, Tanaka, & Drain, 1995) monkeys have less experience with butterflies than humans have with faces and the

Table 3
Statistical results for the rotation conditions, *p* values

Rotation	Effect	J		S	
		RT	Accuracy	RT	Accuracy
15	Rotate figure	0.08	0.01	0.19	0.33
	Rotate tail	0.27	0.31	0.38	0.35
	Interaction	0.15	0.67	0.12	0.38
90	Rotate figure	0.001	0.005	0.02	0.02
	Rotate tail	0.10	0.53	0.45	0.16
	Interaction	0.73	0.38	0.72	0.22

inversion effect has been shown to be less strong for stimulus sets that are less familiar (Diamond & Carey, 1986). This makes it difficult to completely dismiss holistic processing of conjunctive stimuli from these data alone. However, as will be shown in Experiment 2 below, the response time scales with the number of conjunctions required for stimulus recognition and further supports a piecemeal, as opposed to holistic, approach to stimulus recognition. The data do not rule out the possibility of holistic processing by monkeys for highly familiar stimuli, but it does suggest that holistic processing is not the default processing mode for the types of non-face stimuli used routinely in visual classification tasks.

In summary, Experiment 1 confirmed that monkeys can solve XOR tasks using complex, natural, visual images as stimuli. It demonstrated heterogeneity between monkeys and did *not* provide confirmation of holistic processing.

3. Experiment 2

Experiment 1 showed a gap in response times between the XOR and diagnostic stimuli. Was this some sort of threshold effect or was it actually related to the need to detect and report a conjunction? If the latter, then we would hypothesize that there should exist a monotonic relationship between the number of conjunctions required for correctly identifying an image and the accuracy and speed of that determination. But could a monkey even learn a triple conjunction? Experiment 2 taught monkey J a triple conjunction and evaluated the response time and accuracy costs for objects requiring at least one, two, or three diagnostic components.

3.1. Methods

We employed the same basic experimental paradigm, but for stimuli, we adapted the unitization experimental design used by Goldstone (2000) in human studies, but “dressed up” to enhance the visual richness. Individual spline curves (NURBS) were created in 3D Studio Max (Discreet, Montreal, Canada) and “lathed” to create a three dimensional object. A number of these volumes were created and then stacked and fused to create distinguishable abstract bodies (see Fig. 5 for examples).

By design, the objects were created so that 3, 2 or 1 part would be required for identification. Table 4 shows, schematically, the design. The letters represent the parts used to build an object. Item ABC is the top item in Fig. 5, and the next three items are also shown from left to right in the figure. Each unit shares two of its parts with item ABC and so ABC requires identifying something about all three parts for correct identification. However, the other items require only two parts to be identified correctly (e.g., A*Z is a diagnostic pair of parts for object AYZ). In addition, two objects were created with all unique parts (lower case letters in the table). To keep the number of right and left button presses equal for each of the basic parts, there were twice as many ABC trials as for each of the other images.

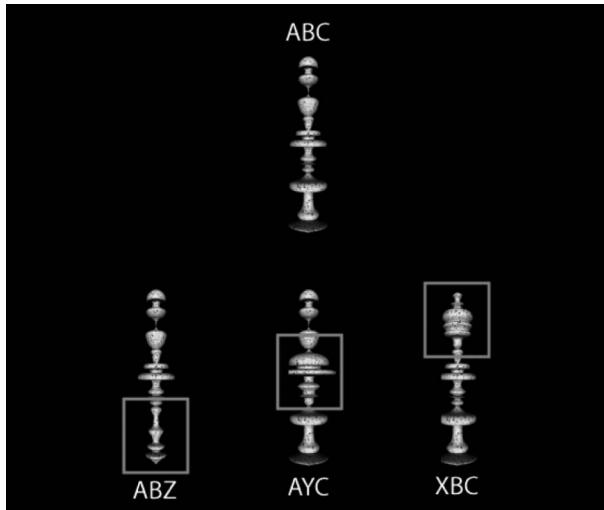


Fig. 5. Four of the seven stimuli used for Experiment 2. The top stimulus required correctly detecting each of its three parts as three other items differed from it in one section only, highlighted here with squares. In addition, there were two figures composed of three unique pieces each and one figure made out from stacking the three highlighted pieces one on top of the other. See also Table 4.

Table 4
Stimulus characterization and response mapping—Experiment 2

Part labels	Response side	Number of parts required for identification
ABC	Left	3
ABZ	Right	2
AYC	Right	2
XBC	Right	2
XYZ	Left	2
def	Left	1
ghi	Right	1

3.2. Results and discussion

This task proved to be very difficult. J underwent 42 separate daily sessions on this task over a span of 133 days. Eventually, J was able to correctly identify all seven stimuli. Unique stimuli were learned earliest (Fig. 6). Interestingly, the response time curves never came together and there appeared to be a “dose” effect with the stimuli requiring the greater number of feature conjunctions showing progressively slower reaction times even after accuracy of response had reached a plateau (permutation test $p < 0.001$).

4. Experiment 3

An interesting issue is why monkeys should easily learn to sort objects along a single dimension but have so much trouble when a second dimension is introduced. If XOR problems are encountered naturally, it might be more usual to occur in a setting where the monkey was already familiar with a number of the basic objects. If this is the case, the laboratory analogue would be to test XOR performance for objects where the animal had already learned the linear, pairwise discriminations. To make this more concrete, consider Fig. 7 which shows half the stimuli used in Experiment 3. To correctly classify each one, it is necessary to consider both the size and shape of the feature on each end of the connecting bar. However, by breaking this set into several smaller ones, as shown in Fig. 8, the monkey can be taught to make the correct button press for each image by attending to only one dimension at a time. Would learning these pairwise discriminations transfer over to a full XOR task?

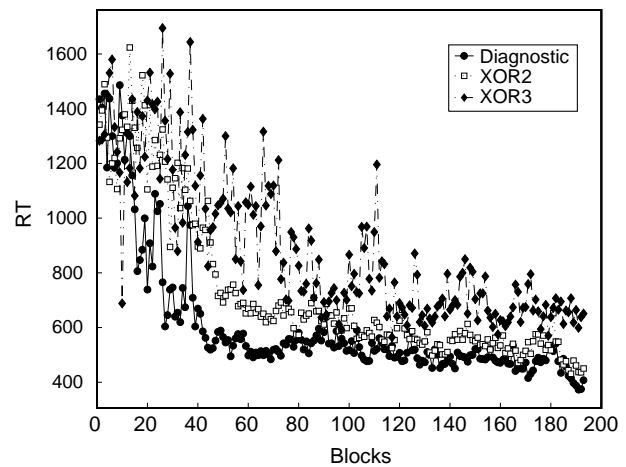
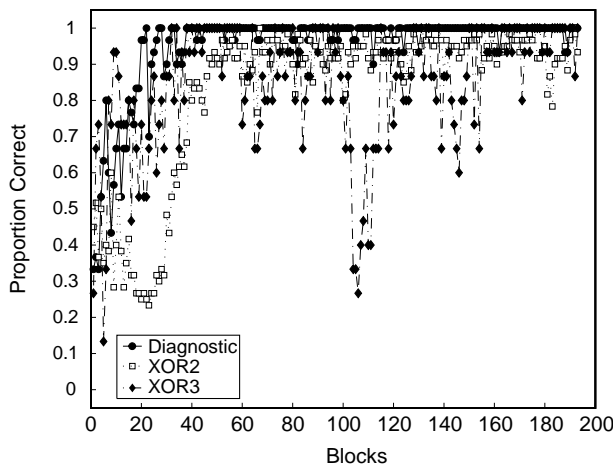


Fig. 6. Learning a triple conjunction. The accuracy and response speed of monkey J is shown for acquisition of a triple conjunction. There were two images that were composed completely of unique parts, four images that required two parts for correct assignment (XOR 2) and one image that required identifying three correct parts for correct assignment (XOR 3). Learning was slow but eventually successful for all images in the set. There appeared to be a cost in terms of response time and accuracy for each additional conjunction required mitigating against the idea that the images were all being processed as wholes or that they had been functionally unitized.



Fig. 7. The set of “near” geometric XOR stimuli. Each stimulus was composed of a triangle connected by a bar to a rectangle. Triangles and rectangles came in small and large sizes. If either both the top and bottom were large, or neither was, the monkey pressed one button. If only one or the other were large the monkey pressed a second button. The separation between top and bottom was 0.15° of visual angle. A second similar “far” set was also used, with separation of 3° of visual angle (Fig. 8).

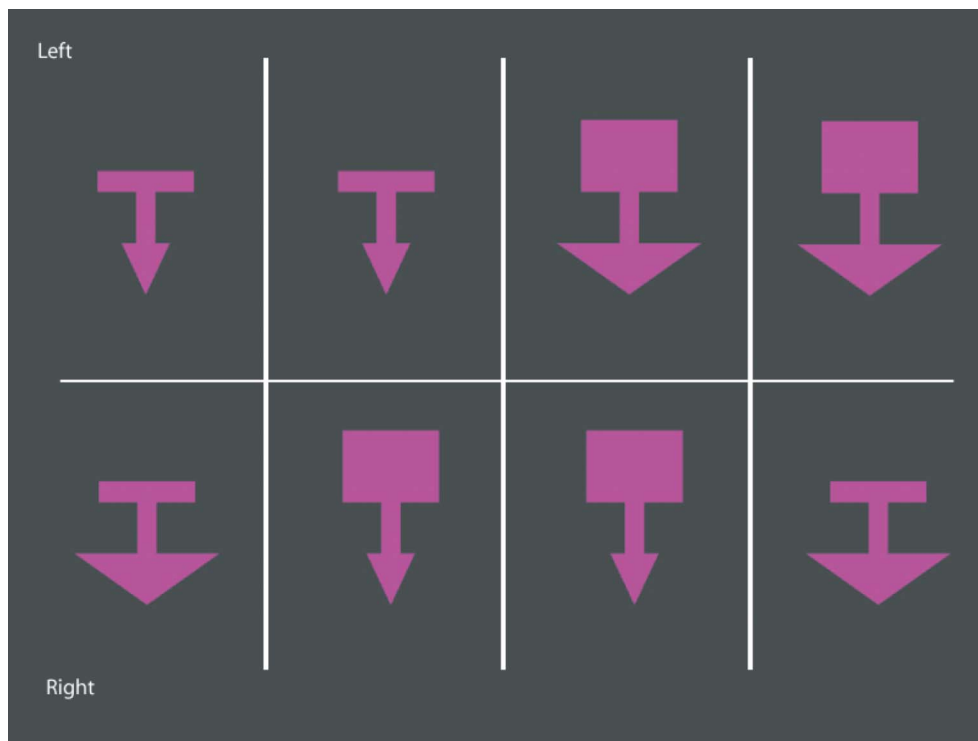


Fig. 8. The set of “far” geometric XOR stimuli in training groups. Sets of training stimuli are separated by white lines. For training blocks, correct selections could be made along a single stimulus dimension (e.g., the triangles in the left most column). During all training and testing experimental blocks, stimuli with short and long connecting bars were mixed.

4.1. Methods

The general testing procedure was the same as for Experiment 1. The stimuli were two-dimensional polygons comprising two triangles and two rectangles which were connected by a bar (Fig. 7). These four parts were shown in

all possible combinations (with the triangle always appearing below the rectangle) to yield four stimuli. In addition, the bar connecting the upper square and lower triangle was either 0.15° of visual angle long or 3.0° of visual angle long. On individual trials, monkey S initiated trials by fixating for 450 ms on a blue fix spot, which was presented in the

center of the monitor. For monkey J, the appearance of the stimulus marked the start of the trial.

For blocks of two-item discriminations in the Training Phase, the monkeys were shown a subset of two of the four objects with the short and long connecting bar versions intermixed for a total of four images (Fig. 8). These four stimuli were repeated 25 times during a single experimental block, to yield a total of 100 trials per experimental block. During the Training Phase, each monkey was given several experimental blocks of one two-item discrimination during a single daily session (different two-item discriminations were never mixed within a daily session). Each monkey was required to meet a criterion of greater than or equal to 90% accuracy for two consecutive blocks during a daily session on the current discrimination. Once the monkey completed this criterion, the following daily session would start with a new two-item discrimination.

Once each monkey completed all four two-item discriminations separately, it was given two daily sessions (a total of 24 experimental blocks) in which the four two-item discriminations were shown for the first time within the same *daily* sessions (Block Phase), but still there was no mixing of discriminations within a particular experimental block. The order of the experimental blocks of two-item discriminations was randomized.

Following the Block Phase, the monkeys proceeded to the XOR Phase. All four stimuli (Fig. 7) in both the short and long connecting bar versions were intermixed in a single experimental block and the monkey was required to use both parts (the triangle and the rectangle) to make the proper discrimination. The four stimuli, shown in both the close and far conditions, were each shown 10 times to yield 80 trials per experimental block (4 images \times 2 closeness conditions \times 10 repetitions = 80 trials).

4.2. Results and discussion

The two monkeys completed a median of 20 experimental blocks for each two-item training phase discriminations (range: 16–57). As a comparison to training with butterflies (Fig. 4), Table 5 shows the number of 15 item analysis blocks completed for each of the four two-item steps for the two monkeys. Both monkeys successfully meet criterion for all four two-item discriminations.

Table 5
Number of 15 trial repetitions

Phase of experiment	Monkey S	Monkey J
Training		
Step 1	25	37
Step 2	33	34
Step 3	32	34
Step 4	32	95
Block Phase	20	20
Intermingled Phase	20	20

Spacing of the relevant features affected task difficulty as the far condition (3.0° visual angle) was somewhat more difficult during this training phase. Analyzing the first 20 blocks of 15-trial repetitions for each of the two-item discriminations, we found a significant difference between the close and far conditions for monkey J (only in the accuracy measure, $p < 0.05$). For monkey S, permutation tests indicated a significant difference between the close and far conditions for both accuracy and response speed ($p < 0.05$).

During the Block Phase, the monkeys were shown all four two-item discriminations within the same daily session. A one-tailed binomial test ($p < 0.05$) indicated that the accuracies for both monkeys were significantly above chance.

In the Intermingled Phase, the full XOR task was administered. There was virtually immediate transfer of performance to the XOR task with performance near ceiling. A one-tailed binomial ($p < 0.05$) indicated that the accuracies for both monkeys were significantly above chance for both the close and far conditions. Permutation tests were also done on the XOR task performance. For monkey J, there was a significant difference between the close and far conditions for response speed only ($p < 0.001$). For monkey S, there was a significant difference for distance in both accuracy and response time ($p < 0.01$).

The main result was that training in a stepwise fashion aided the monkeys in solving the XOR task (Fig. 9). However, in Experiment 3 the effect of training was confounded with stimulus familiarization. That is, as a result of the training phase, each monkey saw each stimulus a large number of times. In keeping with the idea that any potential natural analogue of an XOR task is likely to exist mostly for objects with which the organism already has some familiarity, we can ask if familiarity with objects alone is enough to facilitate acquisition of an XOR task? Stated simply, was it the training that helped in Experiment 3 or simply the hundreds of exposures to each image?

5. Experiment 4

For Experiment 4, we took advantage of the fact that two of the monkeys had extensive familiarity with a set of pocket watch images that were amenable to being modified to create a visual XOR test (Fig. 10). Monkey S had extensive passive exposure to a group of pocket watch images, and monkey T had been over-trained on a classification task involving the same pocket watch pictures. New hybrid stimuli were constructed from these original, familiar watches. By testing on these hybrid pocket watches, we could directly evaluate whether passively or actively gained familiarity aided solving an XOR task.

5.1. Methods

The general behavioral testing scheme remained the same. Each watch was divided into two parts, the circular face and everything outside the face. Parts from watches

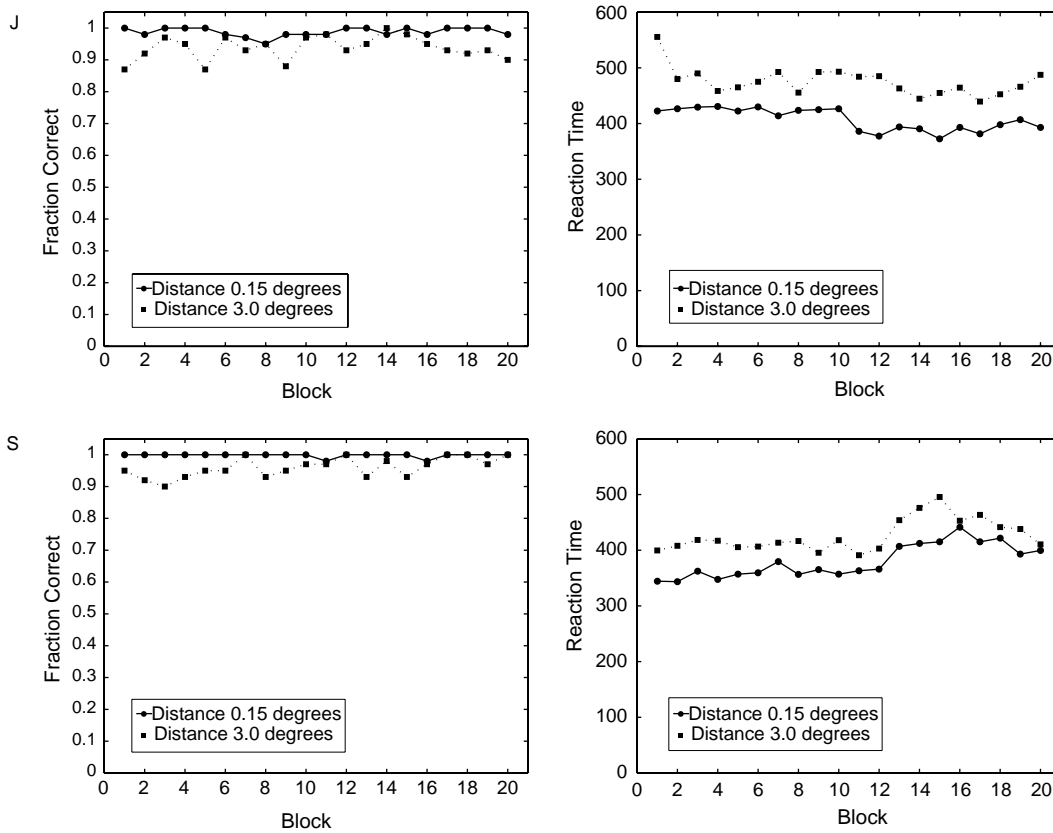


Fig. 9. XOR results for Experiment 3. The accuracy (left) and response speed (right) are shown for monkeys J and S on the XOR test of Experiment 3. As is clearly seen, the preliminary training on each pairwise comparison allowed them to rapidly transition to a full XOR task. Compare the early trials in this experiment with the early trials of Figs. 4 or 11.

that had been trained to the same button were recombined, and the resulting new watches were assigned to the button opposite the sources of their parts. For example, a hybrid formed from a face and a frame belonging to left watches would require a right button press (Fig. 10). Eight hybrids were constructed, four per side of response, and added to the eight familiar watches for a total of sixteen

stimuli. These were all presented eight times in each experimental block (yielding an experimental block size of 128). During early training, these blocks were sometimes intermingled with “refresher” blocks in which the monkey had to classify only the old stimuli (monkey T) or to perform the familiar passive viewing task with the old stimuli (monkey S).

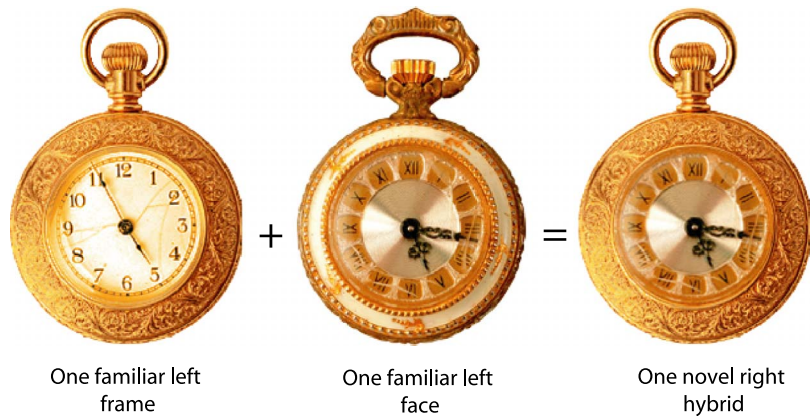


Fig. 10. An example of the construction of the hybrid watch stimuli. Each monkey (S & T) had familiarity with a large collection of pocket watch images. Monkey T had learned to associate each with a button press. From this set of familiar and learned watches we divided the images into “faces” and “frames,” which were recombined using computer image processing software into new, hybrid images. The button press associated with the hybrids were mapped to the opposite hand. In this example two left familiars were combined to make a right hybrid; the opposite pairings occurred equally often. The faces and frames played the same roles as the tail spots and butterfly bodies of Experiment 1 (Fig. 1).

5.2. Results and discussion

Monkey S completed 48 experimental blocks and monkey T 64 (25 and 32 of the 15-trial analysis blocks, respectively). Both were able to learn the XOR task, with monkey S reaching and maintaining 85% accuracy in the 19th analysis block and monkey T in the fifth (Fig. 11). Monkey T responded with better speed and accuracy than did monkey S after reaching this performance level. We used permutation tests to compare the accuracy and response times between hybrid and familiar stimuli, once the task was learned. Monkey S was less accurate ($p < 0.01$) and slower (mean response time difference 48 ms, $p < 0.001$) with hybrid stimuli than with familiar stimuli. The results for monkey T were similar (accuracy, $p < 0.001$; mean response time difference 70 ms, $p < 0.001$).

It is important to highlight that since monkey T had already learned button presses for half the items he only had to learn the XOR items while monkey S, though familiar with half the items, had to learn button press mappings for all sixteen items. This difference is noted in the early performance of T in Fig. 11. He was initially nearly perfect on the learned items and almost uniformly wrong on the new items where the appropriate button press was opposite that predicted from the individual pieces (e.g., left face + left frame = right button press and vice versa).

Both monkeys were able to learn this task with fewer presentations of each stimulus than was required for the XOR task in Experiment 1. This is not due to general expertise with conjunction tasks, as the order of training on the two tasks was reversed for the two monkeys. Monkey S performed Experiments 1 and 3 before being trained on the XOR watch task of Experiment 4, whereas Monkey T achieved high levels of accuracy on the XOR watch task before being trained on the butterfly stimuli of Experiment 1. If preliminary exposure to conjunction type tasks alone was the explanation for performance differences, then mon-

key S should have been superior to monkey T. In addition, before any exposure to the stimuli of Experiment 1, T was performing at greater than 90% accuracy for all hybrid watches except one (84%). After slightly greater number of exposures per hybrid stimulus on the butterfly task (225/image) his performance was uniformly worse (range 39–77%). It appears that training on one half of the conjunction task facilitates learning of the complete task. It also appears that passive exposure alone, without the learning of any discrimination, may also be sufficient to improve performance.

6. General discussion

All our variations of XOR type tasks required considerable training of the monkeys. We did not find that the variety of objects per se, or the physical spacing of the critical elements was of central importance. Rather, it appears that the monkey's familiarity with the actual images is what simplifies XOR task acquisition. This familiarity may be acquired through simple passive viewing of the image set, but it is more effective when the images are the objects of some action or if the individual pairwise discriminations are learned first.

The XOR tasks we employed are highly contrived. However, one circumstance in which we might naturally meet a situation approximating the XOR task is when we are exposed to a set of familiar well learned objects. We would then have to make fine discriminations among and between them, perhaps such as when we need to discriminate faces. It does not seem an unreasonable deduction to think that stimulus familiarity would show a powerful effect on XOR task acquisition.

Whether or not this surmise is accurate, the empirical results were clear that stimulus familiarity was important. This effect was present for passive stimuli exposure (Experiment 4, monkey S), active stimuli learning (Experiment 4,

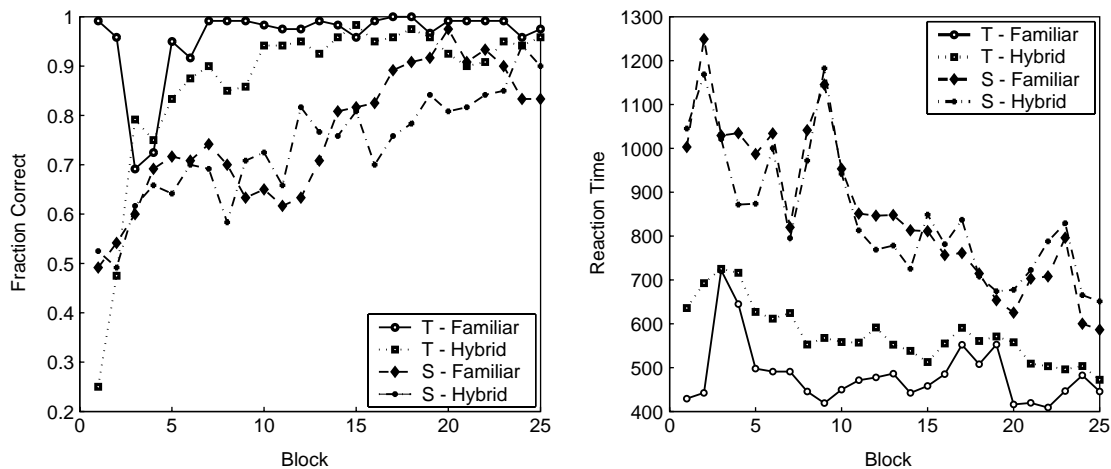


Fig. 11. Performance of monkeys S & T for learning the XOR watches. Both monkeys showed a comparatively easier time learning the pocket watch XOR stimuli (compared to the butterfly task) even though there were more items. The monkey who had actively learned the watches (T) as opposed to passive familiarization (S) appeared to acquire the task more quickly.

monkey T) and, most dramatically, for prior learning of pairwise comparisons (Experiment 3, monkeys J and S). In both of the experiments examining the effects of stimulus familiarity, we employed the entire image/object. Whether familiarization with the relevant image components, e.g., the frames and faces of the pocket watch separately, would be effective was not determined.

We did not select faces as stimuli because they are a category of stimuli for which our laboratory monkeys have had extensive, but unquantifiable, experience. We could not, therefore, determine the effect of familiarity with face stimuli on an XOR task. It seems likely, though, that the monkeys' extensive prior experience with faces could promote their ability to learn XOR/non-linear mappings quickly for this stimulus class.

Our motivation for evaluating XOR type tasks was their potential in examining the neurophysiological correlates of “binding.” Our results lead us to several conclusions regarding their application for this purpose. First, since training on an XOR mapping is so prolonged, it will generally be impractical to select stimuli and train an XOR task in the confines of an individual recording session. It may be possible to train animals in advance on a relatively large collection of stimuli and then rely on chance and persistence to isolate cells responding to an informative subset. Second, even if two cells responding to two distinct image components are identified, and the monkey is capable of solving an XOR task based on these stimuli, definitive results may not be achieved. We failed to produce behavioral evidence that XOR stimuli are being combined into a new “whole.” In all of our experiments, all monkeys took longer to respond to conjunction stimuli than stimuli that could be identified by single diagnostic features. The more conjunctions required, the longer the average response time. Further, manipulating the objects by rotating image features, failed to show effects consistent with configural or holistic processing; which is ultimately what synchronous neuronal firing is supposed to account for. Third, monkeys may differ from humans in their response and approach to XOR tasks. In a recent study by Smith et al. (2004), four macaques were tested on six different learning tasks involving different configurations of the same basic stimuli (differently colored and sized geometric shapes). For a group of human subjects the XOR task was the second easiest, whereas for the monkeys it was the second most difficult. Therefore, monkeys may not approach XOR tasks as humans do. Consistent with this difference, Goldstone suggested the response time curves for humans solving unitization tasks came together with repeated practice. We did not see this result for any of our tasks, including the one (Experiment 2) that most closely duplicated his format. Thus, some circumspection would be necessary in drawing conclusions from XOR tasks as a substrate for binding studies, especially if they yielded negative results. This is particularly unfortunate as, in the absence of a theory of what constitutes a visual “feature,” it may be problematic to produce experimental situations in which a direct test of

the binding by synchrony hypothesis is possible for higher order vision.

References

- Baker, C. I., Behrmann, M., & Olson, C. R. (2002). Impact of learning on representation of parts and wholes in monkey inferotemporal cortex. *Nature Neuroscience*, 5, 1210–1216.
- Baker, C. I., Olson, C. R., & Behrmann, M. (2004). Role of attention and perceptual grouping in visual statistical learning. *Psychological Science*, 15, 460–466.
- Deisig, N., Lachnit, H., Giurfa, M., & Hellstern, F. (2001). Configural olfactory learning in honeybees: negative and positive patterning discrimination. *Learning & Memory*, 8(2), 70–78.
- Diamond, R., & Carey, S. (1986). Why faces are and are not special: An effect of expertise. *Journal of Experimental Psychology: General*, 115, 107–117.
- Farah, M. J., Tanaka, J. W., & Drain, H. M. (1995). What causes the face inversion effect. *Journal of Experimental Psychology: Human Perception and Performance*, 21, 628–634.
- Goldstone, R. L. (2000). Unitization during category learning. *Journal of Experimental Psychology: Human Perception and Performance*, 26, 86–112.
- Keating, C., & Keating, E. (1993). Monkeys and mug shots—cues used by rhesus-monkeys (macaca-mulatta) to recognize a human face. *Journal of Comparative Psychology*, 107, 131–139.
- Kehoe, E. J., & Graham, P. (1988). Summation and configuration: Stimulus compounding and negative patterning in the rabbit. *Journal of Experimental Psychology: Animal Behavior Processes*, 14, 320–333.
- Kehoe, E. J., & Macrae, M. (2002). A Neuroscientist's Guide to Classical Conditioning. In *Fundamental behavioral methods and findings in classical conditioning* (pp. 171–231). New York: Springer.
- Kobatake, E., Wang, G., & Tanaka, K. (1998). Effects of shape-discrimination training on the selectivity of inferotemporal cells in adult monkeys. *Journal of Neurophysiology*, 80(1), 324–330.
- Kumara, H. N., Singh, M., Sharma, A. K., Singh, M., & Kumar, M. A. (2000). Faunal component in the diet of lion-tailed macaque. *Primate Report*, 58, 57–65.
- Lober, K., & Lachnit, H. (2002). Configural learning in human pavlovian conditioning: Acquisition of a biconditional discrimination. *Biological Psychology*, 59(2), 163–168.
- Pesarin, F. (2001). *Multivariate permutation tests with applications in biostatistics*. Chichester: Wiley.
- Redhead, E. S., & Pearce, J. M. (1995). Stimulus salience and negative patterning. *Quarterly Journal of Experimental Psychology B*, 48(1), 67–83.
- Rescorla, R. A., Grau, J. W., & Durlach, P. J. (1985). Analysis of the unique cue in configural discriminations. *Journal of Experimental Psychology: Animal Behavior Process*, 11(3), 356–366.
- Saavedra, M. A. (1975). Pavlovian compound conditioning in the rabbit. *Learning & Motivation*, 6, 314–326.
- Saunders, R. C., & Weiskrantz, L. (1989). The effects of fornix transection and combined fornix transection, mammillary body lesions and hippocampal ablations on object-pair association memory in the rhesus monkey. *Behavioural Brain Research*, 35(2), 85–94.
- Shadlen, M. N., & Movshon, J. A. (1999). Synchrony unbound: A critical evaluation of the temporal binding hypothesis. *Neuron*, 24, 67–77.
- Sigala, N. (2004). Visual categorization and the inferior temporal cortex. *Behavioural Brain Research*, 149(1), 1–7.
- Singer, W. (1999). Neuronal synchrony: A versatile code for the definition of relations? *Neuron*, 24, 49–65.
- Smith, J. D., Minda, J. P., & Washburn, D. A. (2004). Category learning in rhesus monkeys: A study of the shepard, hovland, and jenkins (1961) tasks. *Journal of Experimental Psychology General*, 133, 398–414.
- Thorpe, S. J., O'Regan, K., & Pouget, A. (1989). *Neural networks from models to applications, chapter humans fail on XOR pattern classification problems*. Paris: IDSET.