Journal of Comparative Psychology
2004, Vol. 118, No. 1, 3–13

0735-7036/04/$12.00 DOI: 10.1037/0735-7036.118.1.3

Levels of Abstraction in Orangutan (*Pongo abelii*) Categorization

Jennifer Vonk and Suzanne E. MacDonald
York University

Levels of abstraction have rarely been manipulated in studies of natural concept formation in nonhumans. Isolated examples have indicated that animals, relative to humans, may learn concepts at varying levels of abstraction with differential ease. The ability of 6 orangutans (*Pongo abelii*) of various ages to make natural concept discriminations at 3 levels of abstraction was therefore investigated. The orangutans were rewarded for selecting photos of orangutans instead of humans and other primates (concrete level), primates instead of other animals (intermediate level), and animals instead of nonanimals (abstract level) in a 2-choice touch screen procedure. The results suggest that, like a gorilla (*Gorilla gorilla gorilla*) tested previously (Vonk & MacDonald, 2002), orangutans can learn concepts at each level of abstraction, and unlike other nonhumans, most of these subjects rapidly learned the intermediate level discrimination.

It is generally accepted that animals are able to classify and categorize items in their surroundings. For instance, they are able to distinguish food from nonfood items and predators from prey. The ability to categorize might be expected to evolve in a variety of species because such discriminations are critical to an animal's survival (Huber, 1999; Menzel, 1997; Mervis & Rosch, 1981). However, the investigation of concept formation in nonhumans is still admittedly controversial (Huber, 1999). Even when subjects show transfer of prior learning, or generalization, to novel stimuli, it is not clear on what basis the generalization occurs. For instance, the animals might continue to respond on the basis of perceived features rather than on the basis of an overall concept, for example, "select green" as opposed to "select tree."

To further complicate the picture, natural categories consist of groupings of "infinite or open-ended classes of things that occur in nature" (Schrier, Angarella, & Povar, 1984, p. 564), as opposed to arbitrary experimenter defined categories, and can overlap in the sense that features normally associated with one category can appear in exemplars of another category as well (Herbranson, Fremouw, & Shimp, 1999). In addition, natural concepts involve different levels of abstraction, with the number of correlated features both within and between categories decreasing from the most concrete to the most abstract level (Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976). Rosch et al. (1976) theorized that humans form concepts most easily at the basic level in which objects within the category have more correlated attributes relative to objects within more abstract categories. In addition, exemplars of basic level categories also share fewer attributes with members of other categories relative to exemplars of concrete level categories, thus increasing category distinctiveness at the basic level.

This basic level corresponds to what was considered an intermediate level of abstraction in a study by Roberts and Mazmanian (1988). Concepts were defined according to levels of abstraction in the following manner: The most concrete level required subjects to abstract the concept of "kingfisher" from other birds. At an intermediate level, subjects were required to discriminate between birds and nonbird animals. At the most abstract level, animals were to be distinguished from nonanimals. The degree of abstraction was defined in terms of the breadth of the category to be learned. As concepts become more abstract, the defining characteristics of the exemplars become less tied to readily perceivable features and may reflect a more conceptual understanding of category belongingness (Astley & Wasserman, 1999; Eimas & Quinn, 1994; Gelman, 1988). For instance, the features binding various animal species into a single animal category include the ability to breathe, reproduce, eat, and so forth. These features can not be discerned from two-dimensional images. Correct categorization at this abstract level thus depends on a conceptual representation of the depicted object that incorporates knowledge generalized from, but not extrapolated solely from, the physical features of the stimulus.

There has been ample evidence for concept learning at the concrete level of abstraction in both primates and pigeons; however, the basis for the high levels of correct categorization is as yet undetermined. Early work on the formation of natural concepts in animals showed that pigeons rapidly learned to discriminate between pictures that included trees, water, and humans, and pictures that did not, regardless of whether they were reinforced for pecking the S+ (concept present) or S− (concept absent) images, and they showed reliable transfer to new pictures (Herrnstein, 1979; Herrnstein, Loveland, & Cable, 1976). However, despite the lack of evidence to suggest that simple features were more often associated with one category or the other, the possibility that the

Jennifer Vonk and Suzanne E. MacDonald, Department of Psychology, York University, Toronto, Ontario, Canada.

Jennifer Vonk is now at the Cognitive Evolution Group, University of Louisiana at Lafayette.

This research was supported by a Natural Sciences and Engineering Research Council (NSERC) postgraduate scholarship to Jennifer Vonk and by an NSERC operating grant to Suzanne E. MacDonald.

The cooperation and support of the staff at the Toronto Zoo were greatly appreciated. We thank Bev Carter, Karyn Tunwell, Bridget Burke-Johnson, Andrea Beatson, Rick Vos, Mark Bongelli, David Partington, Des Macguire, and Jackie Craig, without whose assistance this study would not have been possible. We also thank John Turtle for programming assistance and Jon Toth for providing protective equipment.

Correspondence concerning this article should be addressed to Suzanne E. MacDonald, Department of Psychology, York University, Atkinson Faculty of Liberal and Professional Studies, 4700 Keele Street, Toronto, Ontario M3J 1P3, Canada. E-mail: suzmac@yorku.ca

pigeons responded on the basis of features versus abstract concepts could not be completely ruled out (Herrnstein, 1979).

Similar studies, suffering from the same interpretive problems, have tested the ability of various primate species to categorize natural objects such as foods, humans, and members of their own and closely related species (D'Amato & Van Sant, 1988; Fujita, 1987; Jitsumori & Matsuzawa, 1991; Schrier & Brady, 1987; Yoshikubo, 1985). Monkeys often show above chance transfer to novel photos from the previously learned categories. However, as with pigeons, it is not always clear that monkeys acquire the relevant concepts as opposed to attending to distinctive perceptual features present in only one class of stimuli. For instance, Schrier and Brady (1987) modified slides of humans in various ways and found that monkeys (*Macaca mulatta*) were more likely to include scrambled or upside down humans in the human category versus silhouettes or nonhuman primates. Jitsumori and Matsuzawa (1991) also manipulated pictures in various ways and found that monkeys (*Macaca mulatta* and *Macaca cyclopis*) showed some transfer of learning to new pictures of full frontal and rear views of humans and silhouettes but did not show transfer to close-up and far human faces. Both patterns of results suggested that subjects did not learn a concept for human under which slides of humans taken from all possible angles and distances should be included.

When D'Amato and Van Sant (1988) examined the errors made by monkeys (*Cebus apella*) in their study testing for the concept human, they discovered that several monkeys incorrectly responded to a nonhuman slide that included a reddish piece of watermelon. Inspection of the slides revealed that about one third of the human slides involved patches of red, usually in the form of clothing. These findings suggest that the monkeys' performance was mediated, or at least distracted, by specific stimulus features, instead of by a general concept of *human*. It is unclear whether control by specific features is incompatible with a concept process; however, for concept formation to transcend perceptual discrimination learning, there must be certain constraints on the process. "One such constraint is that irrelevant features ought not to exert strong control over the subject's categorization behaviour" (D'Amato & Van Sant, 1988, p. 53). The red patch in the problematic pictures constitutes an irrelevant feature in that it is not a necessary or unique component of the target stimulus or category. Thus, the critical aspect is whether the features attended to are relevant or irrelevant. D'Amato and Van Sant (1988) doubted the conceptual nature of their subjects' classifications given that they continued to classify slides containing red patches as S+ instances, even when they clearly belonged to the S− category.

The human participants in Roberts and Mazmanian's (1988) study readily learned concepts at all three levels of abstraction. However, if animals learn concepts only when exemplars within a category look very much alike, it might be expected that they would learn the most concrete level concept the most readily. In fact, this turned out to be the case. It is surprising, though, that both pigeons and monkeys performed more accurately on the more abstract problem of animals versus nonanimals relative to birds versus other animals. This finding is in contrast to Rosch et al.'s (1976) basic category model that predicts that concept formation becomes more difficult as problems become more abstract. More recently, a study examining levels of abstraction in a young gorilla demonstrated even better transfer to novel photos of abstract category discriminations (Vonk & MacDonald, 2002). The gorilla in this study also learned the concrete and abstract discriminations more easily than the intermediate level discrimination. The utility of distinguishing between levels of abstraction in nonhuman concept formation must therefore be questioned. Perhaps the extent to which features overlap between and within categories is a more reliable measure of which categories will be most readily discriminated. The ease with which S+ and S− exemplars can be discriminated perceptually may not always correspond perfectly to the breadth of the categories to be learned.

Subjects in concept learning experiments are typically shown hundreds of images over hundreds of sessions and rarely show transfer that reaches the same level of performance as that obtained on previously learned stimuli (cf. Vonk & MacDonald, 2002). In these studies, it is difficult to distinguish between true concept learning interpretations and accounts whereby the animals rely on more perceptually based skills. For instance, an individual could learn to respond appropriately because it remembers past associations between responses to specific exemplars and rewards (exemplar theory). A subject may also generalize to new instances on the basis of perceptual similarity by analyzing features rather than concepts (stimulus generalization).

Here a touch screen was used to present a limited number of pairs of photographs to orangutan subjects for one to five sessions a week, thus limiting the number of exemplars that could be used to generalize to the concept being learned. This method reduces the likelihood of stimulus generalization, making it more difficult to reach criterion on that basis alone (D'Amato & Van Sant, 1988). In addition, novel or transfer photographs were not intermixed with familiar stimuli; rather, on transfer sessions, the subjects were shown an entire set of novel stimuli.

Of the ape species, only chimpanzees (Brown & Boysen, 2000; Fujita & Matsuzawa, 1986; Tanaka, 2001) and one gorilla (Vonk & MacDonald, 2002) have been tested in studies of concept learning. Moreover, many of the chimpanzees that have demonstrated considerable success in such experiments were language trained or had had extensive experience with similar experiments (Premack, 1983; Tomasello & Call, 1997). Therefore, it is difficult to determine how other apes, such as orangutans, particularly those with limited or no training in learning concept discriminations, would perform in such tasks.

Method

Subjects

The subjects were 6 zoo-housed Sumatran orangutans (*Pongo abelii*), 3 males (Dinding, Molek, and Dinar, approximately 42, 22, and 13 years of age, respectively) and 3 females (Abby, Puppe, and Jahe, approximately 41, 34, and 2.5 years of age, respectively). Jahe and Abby were experimentally naïve, whereas Puppe and Dinar had participated in a spatial memory–foraging experiment (MacDonald & Agnes, 1999). Dinding and Molek were previously housed at Yerkes Primate Research facility, but neither had previously participated in a computer touch screen or concept formation study.

Materials

The orangutans were first rewarded for selecting photographs of orangutans (S+) when these photographs were paired with human photos (S−). No photo was included in more than one photo set, so that each of the three photo sets included 20 completely novel photos (for a total of 60 photographs, 30 of orangutans and 30 of humans). The photos varied between images of individuals and groups and between close-ups and full body shots. Age, gender, and orientation of the subjects of the photos were also mixed. S− and S+ exemplars within a set were matched on as many of these aspects as possible. All of the photos were in color except for one human and two orangutan photos. Most, but not all, of the humans were Caucasian. We unsystematically varied these factors, as opposed to holding them constant (as in Brown & Boysen, 2000) so that we might be able to detect a pattern in both errors and correct choices (see also Fujita & Matsuzawa, 1986). If the orangutans attended to the relevant features of the stimuli (those that varied between the S+ and S− exemplar sets), then no irrelevant factors should significantly control their responding.

The orangutans were subsequently given the task of discriminating between photos of orangutans (S+) and other members of the primate family (S−). Again there were three photo sets of 20 photos each, half of orangutans and half of nonorangutan primate species. The other primate photos were composed of prosimians, New World monkeys, Old World monkeys, and the other ape species (gorillas, chimpanzees, bonobos, gibbons). Each set contained photos of primates that were not included in the previous set. All of the photos varied along the same dimensions as specified previously, but all were in color.

It was possible that these discriminations could be made on the basis of a single feature, that is, a reddish color. Therefore, the orangutans were given two new sets of photographs (one for initial transfer and training and one for additional transfer). Again these photo sets consisted of 20 photos each and included some black and white photos of orangutans as S+ exemplars, whereas all of the S− exemplars were primates (except for a red panda and a red squirrel) that were colored similarly to orangutans. Because only this feature (reddish coloration) was held constant across all orangutan photos and because this color was so rare among the previous S− category exemplars, it was felt that eliminating the use of this feature as a cue would determine whether the orangutans' responses depended on this single perceptual feature rather than a concept for the species depicted. In the second set of these photos, a photo of a black and white gorilla was also included to determine if the subjects would incorrectly learn that all black and white photos were associated with reward.

Subsequent tests required the subjects to discriminate between increasingly abstract categories; at the intermediate level, between primates (S+) and nonprimate animals (S−), and at the abstract level, between animals (S+) and nonanimals (S−). There were three primate–nonprimate photo sets of 20 photos each and four animal–nonanimal photo sets, also of 20 photos each. The primate category again included primates from all genera, whereas the animal sets included animals from various taxa such as birds, reptiles, mammals, fish, and insects. We also presented two sets of control primate–nonprimate photos. The first set of control photos included 10 photographs of primates and 10 photographs of nonprimates with their facial features occluded. The second set of control photos included 10 primate photos and 10 nonprimate photos that showed only the faces and none of the bodies. It was hoped that use of these control photos would help us to determine which features were important to recognizing primates as a category.

The S+ animal photos varied in terms of age, number, and orientation of subjects and the other previously described dimensions. The S− nonanimal category in the most abstract discrimination consisted of photographs of both background landscapes and inanimate objects, including two photographs of horse and rider statues and a photograph of clay elephants. All photos were in color.

If the subjects were able to correctly select animal photos it could be because they understood the concept *animal* and the fact that this was the S+ category. Alternatively, it could be that they simply responded to animal photographs because they were more similar to other previously reinforced stimuli, relative to the nonanimal photos. Therefore, accurate performance might reflect true concept learning, but it might also reflect stimulus generalization. Thus, the orangutans were next presented with food (S+) versus animal (S−) photos. Foods are considered an abstract category because membership is determined by function, not by physical appearance. Foods are things that are edible, not things that look alike (Bovet & Vauclair, 2001).

Successful performance would demonstrate that the orangutans were able to use an abstract concept even though they would now be selecting stimuli that appeared perceptually quite different from previously reinforced stimuli and simultaneously inhibiting responding to the photos that appeared most similar to stimuli that previously served as S+. Thus, this test would allow us to determine the extent to which the subjects were likely to use similarity to previously reinforced stimuli as a gage of how to respond in novel tasks. In addition, if a concept, as opposed to a series of features, has been learned, reversal discrimination performance should be acquired more rapidly than original discrimination learning. Thus, the subjects should learn not to select photos of animals more rapidly than they learned to select photos of animals in the previous discrimination task. A detailed list of the features associated with the photographs used here appears in Vonk and MacDonald (2002).

Procedure

The experiments were conducted in a holding area where each orangutan was housed in a separate pen, with the exception of Puppe and Jahe who were housed together. A 13-in. Apple touch screen monitor was placed against the front of the cage where the open bars allowed the subjects access to the screen. Because different animals were kept in the holding area on a rotating schedule, each was tested only one to four times a week for one to three sessions per day. The animals were tested at the same time each day, usually after they had been fed. Data collection took approximately 2 years to complete.

The experiment was programmed in Filemaker Pro 3 software and was run on a Macintosh 5300 PowerBook computer. Each session consisted of 10 trials, with each trial consisting of presentation of a pair of photographs. Within each pair was a photograph of an exemplar from the S+ and the S− category. Thus, each session involved the presentation of the same 20 photographs (10 S+ and 10 S−) until criterion was reached, at which point 20 novel photos were presented. Photos were digitally scanned using DeskPaint software. The photographs were presented in random order with the pairing of photographs from each category also randomized. The presentation location was counterbalanced so that half of the correct photos appeared on each side of the screen during each session, and particular exemplars appeared in both locations through the course of the experiment.

The subjects were required to select one photograph from each pair by touching the screen. When the touch screen was activated, the monitor advanced to a blank screen and the subject was rewarded for only correct choices. Rewards consisted of M&Ms, dried fruits and nuts, or other small but highly preferred foods and were offered by hand. After the subject was rewarded for a correct choice, the screen advanced to the next pair of photographs until the 10th pair had been presented. If the subject made an incorrect choice, the screen immediately advanced to the next trial with no timeout period. The intertrial interval varied but was always less than a minute.

The experimenter was positioned to one side of the touch screen that was protected by a covered cart so that she could not see both of the photos on the screen and therefore could not visually cue the subjects to select the correct photo. The subject's choice could be viewed by watching the laptop positioned behind the touch screen. Correct and incorrect choices were recorded automatically by the software and stored for later analysis.

Initially the criterion had been set at 80% correct responding for four consecutive sessions. However, given that the subjects were tested at such variable and often long intervals, and given the number of distractions that were unavoidable in the zoo environment, it became difficult to maintain consistent responding for four consecutive sessions. Thus, the criterion was changed to 80% correct for two consecutive sessions (or for the last two of three previous sessions at the experimenter's discretion if boredom or distraction was thought to be a factor). This change in the criterion occurred after approximately 7 months of testing, and the subjects were at different points in the experiment at the time.

When criterion was reached, or after a maximum of 30 sessions without reaching criterion, the subject was given transfer trials. A transfer session consisted of 10 trials of novel stimuli that depicted exemplars from the same previously discriminated categories. Transfer photo sets also consisted of 20 photos (10 S+ and 10 S−), randomly paired and presented. Presentation of the photo set continued until the subject had once again reached criterion. In most cases a third set of novel photos was presented, allowing the subjects a second opportunity to demonstrate transfer, indicating whether the number of sessions required to reach criterion decreased across subsequent data sets, implying some degree of learning. Order of presentation of the different photo sets within a single discrimination task varied between subjects.

Data Analysis

Because of the large degree of individual differences, Table 1 presents individual results. However, for brevity and with the aim of presenting the more general trends, the analyses were conducted on data collapsed across subjects. Readers interested in the results of individuals are encouraged to contact Jennifer Vonk and Suzanne E. MacDonald for those details.

One sample t tests were conducted to compare transfer performance with chance (50%) within each discrimination. Following convention, only the first two sessions with novel stimuli were taken as a measure of transfer performance.

To determine whether specific attributes of the stimuli exerted control over the subjects' responding on the task, univariate analyses of variance (ANOVAs) on the percentage of trials on which each photograph was selected were conducted. Separate ANOVAs were conducted on S+ and S− exemplars because features that might be expected to increase the likelihood of selection for S+ exemplars would be the same features that might decrease the likelihood of selection for S− exemplars. For a list of independent variables and the manner in which they were coded, see Table 2. Reinforcement history refers to whether the species of animal depicted in the photograph belonged to a previously rewarded (S+) category, whereas presentation history refers to whether the species had been presented previously, regardless of whether it had served as an S+ or S− stimulus. Two additional variables were also considered. Relative size of the animals in the photographs was coded from 10% to 100%, depending on the perceived portion of the photo that was taken up by the image of the subject relative to the background. Number of individuals present in the photo was a straightforward count. Two independent observers coded the photos, and complete agreement was obtained. Only main effects were included in the analyses. For the most abstract discriminations, we analyzed only the features of the animal photos because it was difficult to quantify components of the nonanimal and food photos that might be expected to guide performance in these tasks. By analyzing the role of the stimulus attributes on performance by all subjects taken together, we could look for overall patterns. If the orangutans made their choices on the basis of the concepts being tested, then none of the listed factors should significantly impact their performance, unless those features served to discriminate the S+ from the S− exemplars.

A series of chi-square tests of independence were conducted to determine which of the coded attributes could serve to differentiate the S+ photos from the S− photos with a probability greater than chance. These attributes appear in Table 3. In addition, chi-square tests of independence were conducted to compare the attributes of photos that were selected 70% of the time or more with the attributes of photos that were selected 30% of the time or less, using the coded attributes of the variables described in Table 2. These analyses revealed that the subjects' choices were always influenced by the same variables that differentiated the S+ and S− sets. Thus, they revealed little more than that the orangutans tended to select only S+ photos at high levels and only S− photos at low levels. Therefore, these analyses are not discussed further. Instead, any deviations from this pattern of responding, such as strong tendencies to select particular S− or not to select particular S+ photos, are discussed.

An alpha value with a significance of .05 was set for all statistical tests reported in this article. This fairly liberal alpha value (uncorrected for multiple analyses) was adopted to avoid Type II errors in the identification of possible stimulus factors controlling the subjects' choices.

Results

Orangutans Versus Humans

Figure 1 depicts the average number of sessions required to reach criterion of 80% correct or better performance for two consecutive sessions across subjects for each level of discrimination. Table 1 depicts the number of sessions required to reach criterion for each individual orangutan, as well as their average percentage correct on the first two sessions of each novel photo set.

The orangutans required from two to eight sessions to reach criterion on the first set of photographs. Recall that originally, criterion was set at 80% correct or better performance for four consecutive sessions. Thus, the orangutans received from 2 to 30 sessions on the training set. Abby required the maximum number of sessions because she did not reach criterion. She initially demonstrated a strong location preference, selecting photos that appeared on only one side of the screen. Eventually, after about 30 sessions of each of the first two sets of orangutan versus human photos, she began to select from both sides of the screen and reached criterion on the third set of photos after only 7 sessions. Performance on transfer trials was significantly above chance, $t(23) = 8.56$, $p < .001$, across all subjects.

A univariate ANOVA of performance for each of the S+ photos revealed no significant effects of any of the stimulus attributes, highest $F(2, 52) = 1.68$, $p = .20$. The same analysis of S− photos also failed to reveal any significant effects, highest $F(2, 56) = 0.83$, $p = .44$. This finding indicates that the orangutans were not using particular single features of the photos to make the discrimination.

Overall, none of the S− photos were selected at high rates, including photos of the experimenter. Only one S− photo was selected more than 50% of the time across all subjects. This was a photograph of an adult male human. No distinctive features point to any reason for the relatively high levels of responding to this photograph. Four S+ orangutan photos were selected at low rates across subjects. A photograph of a female orangutan eating was selected only 27% of the time. Again, the reason for this was not obvious. Two other photos that were selected less than 50% of the time were those that were dark and relatively difficult to discern.

Orangutans Versus Other Primates

The orangutans required from 3 to 17 sessions to reach criterion on the first set of these photographs. Overall, performance on transfer trials was significantly above chance, $t(23) = 7.48$, $p < .001$.

A univariate ANOVA of selections of each photo, with the various attributes of the photos as factors, revealed no effects, although the effect of size approached significance for positive

Table 1
Sessions Required to Reach Criterion of 80% or Better for Two Consecutive Sessions and Percentage Correct Responding on First Two Sessions With Novel Stimuli for Each Type of Discrimination for Each Subject

S+	S−	Set	Sessions to criterion	Performance on first two sessions: Session 1	Session 2
Dinding					
Orangutans	Humans	1	2	90	90
		2	2	80	90
		3	3	80	100
Orangutans	Other primates	1	3	70	90
		2	2	90	80
		3	3	63	80
B&W orangutans	Other red primates	1		70	50
		2	2	80	100
Primates	Nonprimates	1	9	70	60
		2	2	80	90
		3	5	70	70
Face-only primates	Face-only nonprimates	1		70	50
Body-only primates	Body-only nonprimates	1	3	70	80
Animals	Nonanimals	1	31	60	50
		2	9	40	50
		3	15	60	50
		4	5	70	70
Foods	Animals	1	5	80	70
		2	6	60	70
Dinar					
Orangutans	Humans	1	4	60	70
		2	7	60	60
		3	3	70	80
Orangutans	Other primates	1	12	70	60
		2	8	70	60
		3	2	90	78
B&W orangutans	Other red primates	1	12	80	60
		2	3	90	80
Primates	Nonprimates	1	4	60	89
		2	13	50	50
		3	9	50	60
Face-only primates	Face-only nonprimates	1	16	40	60
Body-only primates	Body-only nonprimates	1	7	80	50
Animals	Nonanimals	1	27	60	50
		2	2	80	80
		3	4	60	70
		4	6	60	60
Foods	Animals	1	20	70	60
		2	14	50	60
Molek					
Orangutans	Humans	1	3	50	80
		2	12	90	50
		3	3	70	90
Orangutans	Other primates	1	4	60	60
		2	12	60	50
		3	15	70	40
B&W orangutans	Other red primates	1	15	70	60
		2	7	60	70
Primates	Nonprimates	1	3	78	80
		2	2	80	90
		3	5	90	70
Face-only primates	Face-only nonprimates	1	3	70	90
Body-only primates	Body-only nonprimates	1	2	80	90
Animals	Nonanimals	1	7	80	60
		2	4	90	50
		3	3	80	75
		4	5	80	70
Foods	Animals	1	4	70	50
		2	7	40	40

(*table continued*)

Table 1 (*continued*)

S+	S−	Set	Sessions to criterion	Performance on first two sessions: Session 1	Session 2
		Abby			
Orangutans	Humans	1	30+[a]	60	60
		2	29	70	50
		3	7	60	70
Orangutans	Other primates	1	2	80	90
		2	6	70	60
		3	2	80	100
B&W orangutans	Other red primates	1	6	60	70
		2	5	60	70
Primates	Nonprimates	1	4	60	70
		2	7	70	70
		3	15	70	60
Face-only primates	Face-only nonprimates	1	2	80	80
Body-only primates	Body-only nonprimates	1	5	70	70
Animals	Nonanimals	1	3	50	90
		2	3	70	80
		3	2	80	90
		4	13	80	60
Foods	Animals	1	4	90	60
		2	2	90	80
		Puppe			
Orangutans	Humans	1	3	50	90
		2	9	75	67
		3	15	70	67
Orangutans	Other primates	1	15	70	60
		2	9	90	60
		3	2	80	78
B&W orangutans	Other red primates	1	6	60	60
		2	3	80	80
Primates	Nonprimates	1	13	70	60
		2	2	90	80
		3	2	80	80
Face-only primates	Face-only nonprimates	1	8	30	30
Body-only primates	Body-only nonprimates	1	4	60	60
Animals	Nonanimals	1	5	60	80
		2	3	70	90
		3	7	60	80
		4	2	90	80
Foods	Animals	1	21	70	70
		2	6	60	60
		Jahe			
Orangutans	Humans	1	7	70	80
		2	5	63	67
		3	2	80	80
Orangutans	Other primates	1	17	70	60
		2	7	70	60
		3	7	60	80
B&W orangutans	Other red primates	1	13	40	50
		2	14	60	70
Primates	Nonprimates	1	11	70	90
		2	2	80	80
		3	3	90	70
Face-only primates	Face-only nonprimates	1		70	70
Body-only primates	Body-only nonprimates	1	8	70	70
Animals	Nonanimals	1	6	80	70
		2	10	60	80
		3	8	70	60
		4	8	70	70
Foods	Animals	1	9	50	40
		2	6	70	80

Note. S+ = concept present; S− = concept absent; B&W = black and white.
[a]Abby did not meet criterion, so she received the maximum number of sessions (30).

Table 2
Variables Included in Analysis and Coding Scheme for Each Discrimination

Variable	Coding scheme 0	1	2	3	4	5
Reinforcement history	Not reinforced	Reinforced				
Presentation history	Not presented	Presented				
Facial features	Absent	Present				
Specific color	Black/white	Red/orange	Black	Brown/gray	Light	Mixed
General color	Black/white	Color				
Species	Insect	Fish	Reptile/amphibian	Bird	Mammal	
Orientation	Front	90°	180°	Lying on back		
Body in shot	Head only	Half body	Full body	Mixed	Partial	
Sex	Female	Male	Unknown	Mixed		
Age	Infant	Young	Adult	Old adult	Mixed	

exemplars alone, $F(3, 134) = 2.42$, $p = .07$. The subjects were more inclined to select photographs in which the proportion of space taken up by the animal relative to the background was relatively high. There were no S+ photos that were selected less than 50% of the time or S− photos that were selected more than 50% of the time. Thus, no particular photo presented difficulty for the subjects when taken together. Even a photograph depicting an orangutan with a dog was selected 68% of the time overall.

Orangutans Versus Other Red Primates

The orangutans required from 7 to 15 sessions with the training set before they reached criterion. Performance on transfer sessions was significantly above chance, $t(23) = 5.07$, $p < .001$.

Table 3
Attributes Whose Mean Coded Values Differ Significantly Between S+ and S− Exemplars

Discrimination and attribute	χ^2	*df*	*p*
Orangutan/human ($N = 40$)			
Reinforced	32.40	2	< .001
Sex	24.54	3	< .001
Body in shot	8.53	2	< .01
Orientation	11.61	2	.003
Orangutan/other ($N = 60$)			
Seen	42.86	1	< .001
Color	52.50	1	< .001
Reinforced	60.00	1	< .001
Age	15.97	4	.001
Orangutan/red ($N = 40$)			
Seen	8.49	1	.004
Reinforced	40.00	1	< .001
No. of individuals	4.33	1	.04
Age	13.59	4	.01
Color	8.53	1	.003
Sex	10.83	3	.01
Primate/other ($N = 60$)			
Reinforced	13.47	1	< .001
Seen	32.31	1	< .001
Color	18.77	4	.001
Body in shot	6.09	2	.05
Age	8.39	3	.04
Species	13.47	4	.01

Note. Seen = presentation history; Reinforced = reinforcement history.

Critically, univariate ANOVAs of selections of each photo revealed that color was not an important variable when collapsed across all subjects. However, the orangutans' selections of S− exemplars was affected by the size of the subject of the photo, $F(4, 94) = 3.06$, $p = .02$, and the part of the body depicted in the photo, $F(2, 94) = 3.04$, $p = .05$. The orangutans tended to select photos in which the individual was larger and in which only the head or part of the body was shown. None of the attributes affected selections of S+ exemplars, highest $F(2, 84) = 2.41$, $p = .10$. These results suggest that previous performance on the orangutan–other primate discrimination was not dependent on color as a cue, even though performance declined when this feature was made unreliable, and did not depend on associations to previously non-reinforced species.

Only one of the S+ photos was selected less than 50% of the time. This was a black and white photo of a mother orangutan and her offspring that was selected 49% of the time. Again, there was nothing distinctive about this photograph that made it obvious as to why the subjects might have selected it at relatively low levels. None of the S− photos were selected more than 50% of the time, including the black and white gorilla photo, which was selected only 24% of the time. The most commonly selected S− photos were of a lemur and a uakari (49%).

Primates Versus Nonprimate Animals

The orangutans required from 3 to 13 sessions with the training set of primate photos before they reached criterion. Performance on transfer trials (including face-only and body-only control photographs) was significantly above chance overall, $t(43) = 8.74$, $p < .001$.

Dinar was tested on the control photos months later and did not show immediate transfer. He required 16 sessions to reach criterion on the faces-only photos and 8 sessions on the body-only photos. Unfortunately, Dinar was beginning to show signs of disinterest in the experiment when this test began and either touched the center of the screen with his fist without making a clear choice or touched only one side of the screen without looking at the photographs. The experimenter thus felt that his performance did not reflect an inability to learn the discrimination but rather a loss of interest in performing the task. On days when Dinar touched the center of specific photos, he performed well above chance.

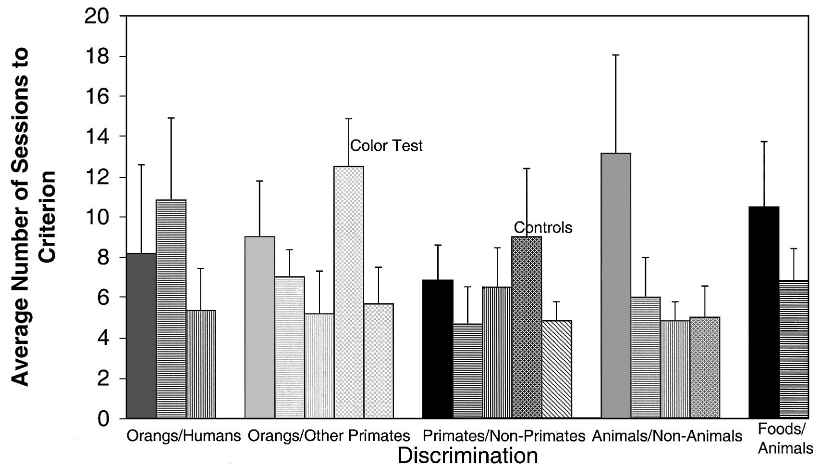

Figure 1. Average number of trials required to reach criterion of 80% correct for two consecutive trials, averaged across all 6 subjects. Each bar depicts performance on a single photo set, along with standard error bars, with photo sets presented from left to right in order of presentation and grouped by level of discrimination. For the discrimination between orangutans and other primates, the last two bars are for photo sets of black and white orangutans versus other red primates (as indicated by "Color Test"). For the discrimination between primates and nonprimates, the last two bars are for photo sets of face-only primates versus face-only nonprimates and body-only primates versus body-only nonprimates, respectively (as indicated by "Controls"). Orangs = orangutans.

Unfortunately Puppe and Jahe became unavailable for testing before they completed the primate control sessions; thus, data are not complete for both sessions in which animal faces were occluded and in which only faces were shown.

An ANOVA of responses to S+ photos with the various attributes of the stimuli included as factors revealed that the only stimulus attribute that significantly affected selections was the number of individuals in the photograph, $F(2, 157) = 3.73$, $p < .03$. The subjects tended to select photos of one or several individuals more often than they selected photos of two individuals. None of the variables significantly influenced selection of S− photos, highest $F(1, 156) = 2.67$, $p = .10$.

Two of the S+ primate photos were chosen at relatively low rates, a photo of three owl monkeys (51%) and a photo of an infant spider monkey (37%). Again, there were no particularly obvious reasons that the subjects might have encountered difficulty with these photos, except that, in the case of the latter photo, they may have greater difficulty recognizing infants and juveniles as members of a species compared to adults. One S− photo was chosen at relatively high rates, a close-up photo of a mouse (52%). Notably, rodents share many physical features with some of the smaller prosimian members of the primate order.

Animals Versus Nonanimals

Performance on this training set varied widely between the subjects. They required from 3 to 31 sessions with the training photographs before criterion was achieved. Dinding was the first to be tested on this discrimination. Initially he was given a set of photos in which the S− exemplars consisted of a range of inanimate objects, including objects such as statues and sculptures. He did not reach criterion after 25 sessions with these photos; thus, he was then shown a set of photos in which the S− category exemplars were of a few familiar objects, such as sticks and landscapes. He was able to reach criterion after only 7 sessions with this new set of photos and then showed above chance transfer to a similar set of photos, on which he reached criterion after only 4 sessions. After reaching criterion on these two sets of photos, he then transferred above 80% back to the first two sets of photos in which the S− exemplars represented a mixture of inanimate objects. Because Dinding had trouble with the first two sets of photos, the other subjects were given the ostensibly easier sets of animal photos first (with the S− category exemplars being restricted to landscapes and the occasional familiar object). Dinar continued to be inattentive in the experiment and required 27 sessions to reach criterion on the first set of photos. However, he did show transfer at 80% correct to the second set of photos and rapidly reached criterion on the next two sets of photos. Overall, performance on transfer trials was significantly above chance, $t(35) = 9.53$, $p < .001$.

An ANOVA of selections with individual S+ photographs with the various stimulus attributes as factors revealed no significant factors, but the presence or absence of facial features approached significance, $F(2, 216) = 2.41$, $p = .09$. No S+ or S− photos were selected less than or more than 50% of the time, respectively.

Foods Versus Animals

Four of the orangutans rapidly reached criterion on this new discrimination (the range for all 6 subjects was 4 to 21 sessions). Taken together, *t* tests showed that transfer was above chance, $t(11) = 2.97$, $p < .01$. Furthermore, most subjects required fewer sessions to reach criterion on this discrimination than on the previous animal–nonanimal tests and did not show below chance performance on initial trials, indicating that they were not simply biased to continue selecting photos most similar to those that had been previously reinforced.

In fact, analysis of the stimuli attributes of the S− exemplars supports this conclusion. An ANOVA of selections with each of the animal photos with the various coded attributes as factors revealed that none of the attributes, including prior reinforcement and presentation history, significantly affected performance, highest $F(1, 101) = 1.93$, $p = .17$. No S+ or S− photos were selected less than or more than 50% of the time, respectively.

Discussion

Results from this experiment revealed significant individual differences, and therefore it is difficult to make general conclusions. Transfer tests are typically used to determine whether subjects have learned concepts or are simply responding to the physical properties of the original stimuli. When a concept has been mastered, performance with the new stimuli should begin at high levels. Although both training and transfer performance on these discriminations was consistently above chance, transfer was not equivalent to criterion performance for all subjects across all discriminations.

There are several theories that suggest that categorization can occur without inferring the formation of abstract concepts (see Pearce, 1994). Most of these theories emphasize the role of physical similarities between exemplars of a category. Feature theory states that categorization depends on the strength of connections. Relevant features belonging to the S+ category will always be reinforced and will acquire associative strength. Irrelevant features, on the other hand, will be present on both S+ and S− trials and will thus gain associative strength only erratically. Relevant features will therefore acquire greater associative strength than irrelevant features. This theory might explain the responding of D'Amato and Van Sant's (1988) monkeys to red patches on the nonhuman slides. However, in the current experiments, orangutan subjects did not seem to be consistently distracted by particular physical features of the stimuli. They appeared to demonstrate an understanding of the concept *primate*. Criterion could not be reached on this discrimination solely through the use of single features because there were no features that were common to all of the primate photos and uncommon to all of the nonprimate photos. In addition, the subjects selected photos of primates when the facial features were occluded and also when only facial features were shown, indicating that their discriminations were not based on attending to facial features or body structure alone. In addition, when a single feature, such as reddish color in the orangutan category, could be used to mediate accurate performance, the data did not support such a strategy. However, the data neither confirm nor deny the possibility that the orangutans attended to various combinations of features.

Exemplar theory states that categorization depends on the ability to remember each instance and the category to which it belongs (Pearce, 1994). In other words, the animal forms simple stimulus–reward associations. Transfer to novel photos is explained by stimulus generalization: the generalization of excitation to new stimuli that resemble those that have previously been rewarded. However, the abilities of these orangutans to demonstrate transfer on trials in which stimuli were widely variant or in which very few prior trials had been presented suggests that exemplar theory is insufficient to explain transfer in these experiments.

Prototype theory states that the animal forms a summary representation that corresponds to the average instance of all the presented exemplars. Once formed, this prototype is activated whenever an exemplar is presented, and responding is appropriate to whichever category matches the prototype (Pearce, 1994). In contrast, concept formation implies that successful categorization does not depend on physical similarity among members of a category. In the current series of experiments, the orangutans were not more likely to select novel exemplars that more closely resembled the majority of exemplars within a category. For instance, they were just as accurate at selecting photos of reptiles or fish as they were at selecting photos of primates or other mammals when presented with animal photos. They were able to correctly select several atypical members of learned categories, such as indris and marmosets in the primate category, and butterflies and worm lizards in the animal category.

Transfer to novel primate and animal photos was particularly impressive because not only were the photos themselves novel, but each subsequent photo set included photos of species that had not been previously seen, and the species that had been previously seen or reinforced were not selected more often. In addition, two subjects who were not available for testing at regular intervals were presented the primate–nonprimate photos without having been previously presented any primate photos. They were able to reach 80% correct performance after only three sessions and despite an interval of several months between testing sessions.

It was possible to look at which cues were being used by the orangutans to solve the concept problems. For instance, the most influential feature in analyses of the orangutans' performance with animal photos was the presence or absence of facial features, the feature that would be most relevant to correctly solving the discrimination. This finding suggests that the features that are relevant to humans may also be relevant to orangutans.

Furthermore, by analyzing the pattern of their errors, we could determine whether these subjects strictly chose items that were the most familiar or if they predominantly selected particular views. Rather than presenting only close-ups of animal faces, mixtures of close-ups and body shots, as well as frontal and side views and individuals as well as groups, were presented. This technique allowed measurement of the orangutans' preferences for particular types of photos, indicated by whether these features were used to make choices, and enabled us to determine whether they made use of features and concepts similarly or differently to the way that humans might. More often than not, the variables that we considered did not significantly impact on the performance of these orangutans. However, some of the orangutans occasionally did seem distracted by these features or performed inaccurately even when not influenced by the variables that we analyzed. Although it seems likely that the orangutans were learning concepts and not simply attending to the perceptual features of the stimuli, it is also possible that their concepts did not perfectly match our own. Less than perfect transfer might be expected if the subjects noticed a change in stimuli and were unsure as to whether the rules had changed. Therefore, some generalization decrement is not altogether unexpected. Despite this, there were instances in which the subjects performed at 80% or better on the very first session with novel stimuli.

It is also important to note that the subjects received only one or two sessions a day and often a week apart, so that transfer sessions sometimes occurred several days after criterion had been reached. This inescapable procedural artifact made it challenging to obtain criterion or significant transfer. In addition, these subjects were not food deprived or without any other form of enrichment. Although they almost always chose to participate in the experiments, there were days when human interaction and the opportunity to manipulate the equipment dominated their attention. Therefore, occasional poor performance may not accurately reflect a lack of ability on their part to perform these tasks on a conceptual basis.

Our group of subjects was unique in that they ranged in age from very young (2 years) to relatively old (42 years). Therefore, it was possible to demonstrate that orangutans, even at very young ages, show some evidence for concepts at various levels of abstraction. The 2-year-old orangutan, Jahe, showed evidence for both intermediate and abstract level concepts. Thus, conceptual abilities may develop early in orangutans.

Conclusion

Until recently, Roberts and Mazmanian (1988) had conducted the only study with nonhuman subjects to manipulate levels of abstraction, an idea that is analogous to the conceptualization of subordinate and superordinate categories investigated by researchers studying concept formation in human children (Mervis & Rosch, 1981; Rosch et al., 1976). Roberts and Mazmanian found that pigeons and squirrel monkeys were able to learn concrete or highly abstract concept discriminations but had difficulty learning discriminations at an intermediate level of abstraction. A similar study showed the same pattern with a young gorilla, although the gorilla showed much superior transfer than monkeys and pigeons at all levels of abstraction (Vonk & MacDonald, 2002). The current findings suggest that orangutans are able to learn concrete, intermediate, and abstract level concepts but have slightly more difficulty with the most abstract concept discriminations. If subjects rely on stimulus features to make their discriminations, it might be expected that concrete and highly abstract categories would be easier to discriminate than intermediate category discriminations in which members of the S+ and S− categories share more features in common. Alternatively, if subjects use concepts to mediate their performance on these tasks, one would expect that the most abstract discriminations would be the most difficult to make. Therefore, the present results and previous research suggest that orangutans may be more likely than monkeys or pigeons, and perhaps even gorillas, to rely on concepts versus features when making category discriminations.

Possibly, it is more adaptive to distinguish between species belonging to more general categories, for example, birds versus fish, as opposed to, for example, blue jays versus cardinals. Therefore, the orangutans' apparent ease in categorizing primates together, relative to orangutans separately from other primates, may not be so surprising, despite the degree of abstractness of the discriminations. The intermediate level discrimination may be more akin, for instance, to making such ecologically relevant discriminations as distinguishing between carnivores and herbivores.

The finding that performance on the concrete level discriminations was not always superior to that on the more abstract problems for all of the subjects could also be an artifact of the order in which the various discriminations were tested. Because the concrete level discriminations were the first problems presented, the subjects had to learn the demands of the task simultaneously. Some improvement might be logically expected over time. Ideally, the discriminations should be presented in random order; however, there were large individual differences, and not all subjects performed the experiment from the beginning. Given the small sample size, it would have been difficult and inappropriate to draw conclusions regarding the relative difficulty of various discriminations with a single subject placed in each ordering condition. Firm conclusions cannot be drawn regarding the relative difficulty of learning concepts at varying degrees of abstraction. Although these subjects appeared to encounter greater difficulty with the most abstract discrimination, relative to the intermediate level discrimination, despite the fact that the latter task occurred earlier in the learning process, it is impossible to determine whether learning of the earlier discrimination interfered with or facilitated learning of the most abstract discrimination. What can be determined is that both gorillas (Vonk & MacDonald, 2002) and orangutans demonstrate above chance transfer to novel stimuli representing discriminations at concrete, intermediate, and abstract levels. They also tend to reach criterion in fewer sessions and show a higher degree of transfer than do other nonhumans previously tested in similar paradigms. Future research should test the order of presentation of various levels of concept discriminations and attempt to further elucidate the features used by various species in performing these tasks.

References

Astley, S. L., & Wasserman, E. A. (1999). Superordinate category formation in pigeons: Association with a common delay or probability of food reinforcement makes perceptually dissimilar stimuli functionally equivalent. *Journal of Experimental Psychology: Animal Behavior Processes, 25,* 415–432.

Bovet, D., & Vauclair, J. (2001). Judgment of conceptual identity in monkeys. *Psychonomic Bulletin & Review, 8,* 470–475.

Brown, D. A., & Boysen, S. T. (2000). Spontaneous discrimination of natural stimuli by chimpanzees (*Pan troglodytes*). *Journal of Comparative Psychology, 114,* 392–400.

D'Amato, M. R., & Van Sant, P. (1988). The person concept in monkeys (*Cebus apella*). *Journal of Experimental Psychology: Animal Behavior Processes, 14,* 43–55.

Eimas, P. E., & Quinn, P. C. (1994). Studies on the formation of perceptually based basic-level categories in young infants. *Child Development, 65,* 903–917.

Fujita, K. (1987). Species recognition by five macaque monkeys. *Primates, 28,* 353–366.

Fujita, K., & Matsuzawa, T. (1986). A new procedure to study the perceptual world of animals with sensory reinforcement: Recognition of humans by a chimpanzee. *Primates, 27,* 283–291.

Gelman, S. A. (1988). Children's expectations concerning natural kind categories. *Human Development, 31,* 28–34.

Herbranson, W. T., Fremouw, T., & Shimp, C. P. (1999). The randomization procedure in the study of categorization of multi-dimensional stimuli by pigeons. *Journal of Experimental Psychology: Animal Behavior Processes, 25,* 113–125.

Herrnstein, R. J. (1979). Acquisition, generalization, and discrimination reversal of a natural concept. *Journal of Experimental Psychology: Animal Behavior Processes, 5,* 116–129.

Herrnstein, R. J., Loveland, D. H., & Cable, C. (1976). Natural concepts in

pigeons. *Journal of Experimental Psychology: Animal Behavior Processes, 2,* 285–302.
Huber, L. (1999). Generic perception: Open-ended categorization of natural classes. *Current Psychology of Cognition, 18,* 845–887.
Jitsumori, M., & Matsuzawa, T. (1991). Picture perception in monkeys and pigeons: Transfer of rightside-up versus upside-down discrimination of photographic objects across conceptual categories. *Primates, 32,* 473–482.
MacDonald, S. E., & Agnes, M. M. (1999). Orangutan (*Pongo pygmaeus abelii*) spatial memory and behavior in a foraging task. *Journal of Comparative Psychology, 113,* 213–217.
Menzel, C. R. (1997). Primates' knowledge of their natural habitat: As indicated in foraging. In R. W. Bryne & A. Whiten (Eds.), *Machiavellian intelligence 11: Extensions and evaluations* (pp. 207–239). Cambridge, England: Cambridge University Press.
Mervis, C. B., & Rosch, E. (1981). Categorization of natural objects. *Annual Review of Psychology, 32,* 89–115.
Pearce, J. M. (1994). Discrimination and categorization. In N. J. Mackintosh (Ed.), *Animal learning and cognition* (pp. 109–134). San Diego, CA: Academic Press.
Premack, D. (1983). The codes of man and beasts. *Behavioral and Brain Sciences, 6,* 125–167.
Roberts, W. A., & Mazmanian, D. S. (1988). Concept learning at different levels of abstraction by pigeons, monkeys, and people. *Journal of Experimental Psychology: Animal Behavior Processes, 14,* 247–260.
Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology, 8,* 382–439.
Schrier, A. M., Angarella, R., & Povar, M. L. (1984). Studies of concept formation by stump-tailed monkeys: Concepts, humans, and Letter A. *Journal of Experimental Psychology: Animal Behavior Processes, 10,* 564–584.
Schrier, A. M., & Brady, P. M. (1987). Categorization of natural stimuli by monkeys (*Macaca mulatta*): Effects of stimulus set size and modification of exemplars. *Journal of Experimental Psychology: Animal Behavior Processes, 13,* 136–143.
Tanaka, M. (2001). Discrimination and categorization of photographs of natural objects by chimpanzees (*Pan troglodytes*). *Animal Cognition, 4,* 201–211.
Tomasello, M., & Call, J. (1997). *Primate cognition*. Oxford, England: Oxford University Press.
Vonk, J., & MacDonald, S. E. (2002). Natural concept formation in a juvenile gorilla (*Gorilla gorilla gorilla*) at three levels of abstraction. *Journal of the Experimental Analysis of Behaviour, 78,* 315–332.
Yoshikubo, S. (1985). Species discrimination and concept formation by rhesus monkeys (*Macaca mulatta*). *Primates, 26,* 285–299.

Received June 24, 2002
Revision received May 29, 2003
Accepted June 2, 2003 ■