Children's Acceptance and Use of Unexpected Category Labels to Draw Non-Obvious Inferences

Vikram K. Jaswal (jaswal@psych.stanford.edu)
Department of Psychology, Jordan Hall, Bldg. 420, Stanford University
Stanford, CA 94305-2130 USA

Ellen M. Markman (markman@psych.stanford.edu)
Department of Psychology, Jordan Hall, Bldg. 420, Stanford University
Stanford, CA 94305-2130 USA

Abstract

Language provides an efficient, uniquely human way of transmitting non-obvious category information between individuals and across generations. To explore whether it can serve this purpose even for very young children, we conducted two experiments: one with 24-month-olds, and the other with preschoolers. Children made non-obvious inferences about perceptually misleading animals. Those who heard the animals called by counter-intuitive category labels made inferences different from those who did not hear the labels, demonstrating an important influence of language on thought, even at 24 months of age. However, preschoolers appear to have been less influenced than toddlers, suggesting that there are limits on children's willingness to accept anomalous category labels.

Introduction

Categorization is fundamental to human cognition, enabling communication and serving as the basis for the representation of objects and for predicting and explaining their behavior (Anglin, 1977; Markman, 1989). Children as young as 3.5 months (Eimas & Quinn, 1994), adults (Rosch et al., 1976), and non-human animals (e.g., Freedman et al., 2001) alike readily use perceptual similarity to determine the category to which something belongs. But when reasoning about an object, or explaining or predicting its behavior, perceptual appearance is not always criterial of category membership. For example, even though eels look like snakes, in order to more accurately characterize their ancestry, behavior and physiological processes, experts categorize them as fish.

Given sufficient experience, children as young as 30 months can form non-obvious categories by noting causal (Gopnik & Sobel, 2000) or functional (Kemler Nelson et al., 2000) regularities between objects. Under some circumstances, even non-human animals can learn to categorize objects in perceptually non-obvious ways (Herrnstein & DeVilliers, 1980). However, recognizing non-obvious similarities can be a slow and laborious process, often requiring experience that is difficult to obtain. Moreover, it requires every individual in every generation to have the experience for him or herself (Tomasello, 1999). Another, arguably more reliable and efficient way to obtain non-obvious category information is through language (Gelman et al., 2000): When a trusted source uses an unexpected category label for an object, it reflects a particular perspective that others have found useful when thinking and reasoning about that object in the past, and it can cause us to revise a classification immediately. For language to have this effect, however, listeners may have to give up a compelling, perceptually based classification in favor of a classification they do not immediately understand.

As adults, we can accept linguistically provided non-obvious classifications (e.g., a whale is a mammal, not a fish) because we implicitly assume that something deeper than surface similarity binds category members together (Medin & Ortony, 1989). Such an essentialist assumption would facilitate the rapid uptake of non-obvious category terms. However, it is not clear whether very young children also expect categories and category terms to encode more than surface similarities. For example, it has been argued that children begin with or quickly develop an expectation that category labels encode similarly shaped objects (e.g., Imai, Gentner, & Uchida, 1994; Smith, 1999). Furthermore, whereas children readily learn words for basic-level categories (e.g., "table"), they have difficulty learning words for superordinate categories (e.g., "furniture"), which have fewer perceptual features in common than basic-level ones (e.g., Horton & Markman, 1980; Mervis & Rosch, 1981; Rosch et al., 1976). Finally, young children sometimes object to parental attempts to correct perceptually based categorization errors (Mervis, 1987).

In two studies, we investigated children's willingness to accept perceptually counter-intuitive classifications on the basis of linguistically provided information alone. In our first study, we asked this question of 24-month-olds. Gelman and her colleagues have found that children as young as 32 months can make inferences on the basis of linguistic information rather than perceptual appearance (Gelman & Coley, 1990; Gelman & Markman, 1986), but their procedure involves memory demands, verbal responses, and linguistic comprehension abilities beyond those of younger children. Our procedure uses an imitation paradigm that minimizes these task demands (see also Baldwin, Markman, & Melartin, 1993; Mandler & McDonough, 1996; Welder & Graham, 2001).

Although preschoolers have been shown to defer to adult labels in the face of perceptually inconsistent information (Gelman & Coley, 1990; Gelman & Markman, 1986), there must be limits on this process. Surprisingly, no one has considered situations under which children reject linguistically provided category information. Using the same paradigm and materials as in the first study, our second study takes up this question with preschoolers.

Experiment 1

Experiment 1 was designed to test whether 24-month-olds would use an experimenter-provided, sometimes counter-intuitive, category label when making non-obvious inferences. In particular, we were interested in whether they would use the label to make an inference that was different from the one they would make without a label.

Method

Participants Participants were 32 toddlers, ranging in age from 23 months, 6 days to 25 months, 28 days ($\underline{M}$ = 24 months, 10 days). Five additional toddlers were tested, but their data are not included due to extreme fussiness resulting in an inability to complete at least half of the session (4), or due to experimenter error (1).

Stimuli Eight animals from familiar categories were grouped into four sets based on similar sizes and body shapes: Cat-dog, horse-cow, bear-pig, and squirrel-rabbit. Category familiarity was confirmed by consulting the MacArthur Communicative Development Inventory for toddlers (Dale & Fenson, 1996), which indicated that at least 44% of 24-month-old children could produce the word associated with each animal (range: 43.9% for "squirrel" to 91.6% for "dog").

Realistic, color drawings of a typical exemplar of each animal were obtained from commercially available picturebooks, and were digitized for computer manipulation (hereafter, "standard animals"). In addition, three test animals for each set were generated on Adobe PhotoShop from the standard animals of that set: Two were typical exemplars (hereafter "typical test animals"), and were created by manipulating the coloration of the standards; the third test animal looked more like one standard animal than the other (hereafter, "misleading test animal"). Misleading test animals were created by using one of the standards (or an additional typical image) as a base, and adding features from the other standard image of that pair. Examples of standard and misleading animals from the cat-dog set are shown in Figure 1.

Sixteen graduate students rated each animal, including the standards, from each set on a 7-point continuum. At one end of the continuum was the category label for one of the animals in a given set (e.g., "cat"), and at the other end was the category label for the other animal in that set (e.g., "dog"). Subjects indicated on the continuum the position of each animal image, with 4 being ambiguous—exactly

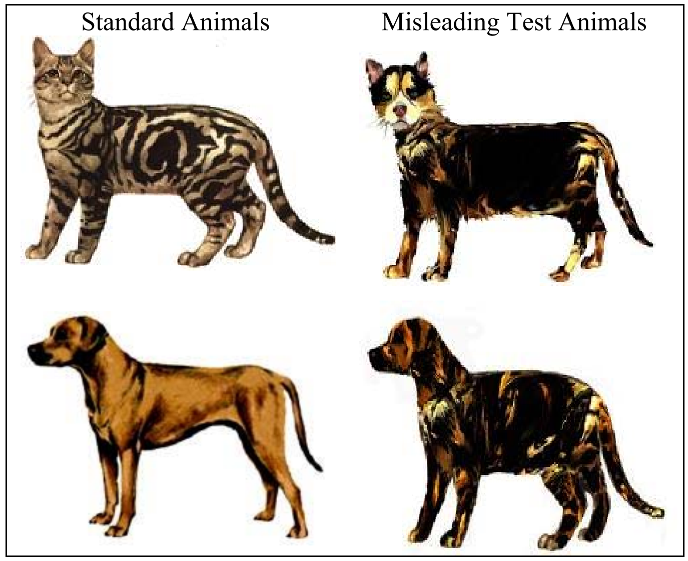

Figure 1: Sample stimuli from the cat-dog set. Each stimulus set was made up of two standard animals, two typical animals (not shown, but very similar to the standards), and one of two misleading animals.

halfway between the two. The average ratings of the standards and typical test animals were always at the ends of the continuum (between 1 and 2 or between 6 and 7), and those of the misleading test animals were always slightly closer to the center, but leaning in one direction (between 2 and 3, or between 5 and 6) and significantly different from 4 and from each other. A full display of the stimuli and details of their adult ratings can be obtained at http://www-psych.stanford.edu/~jaswal/Stimuli/index.html.

All images were sized to approximately 3 to 4 inches wide, with their heights constrained by their widths. Each image (and its left-right reverse) was printed on a 600 dpi color printer, cut out, laminated, and mounted into a small stand that allowed it to remain upright.

Each set of animals was associated with a particular activity, and members of the same pair were associated with contrasting props or dioramas (hereafter, "props"). Activities were chosen that could be easily and clearly demonstrated with a 3-dimensional prop, and that might (though not necessarily) be familiar to children: The cat played with a ball of yarn, the dog played with a stick; the horse slept on the hay, the cow slept on the grass; the bear lived in the forest, the pig lived in the mud; the squirrel ate a nut, the rabbit ate a carrot. Props were purchased or specially constructed.

Design The 32 children were randomly assigned to a label or a no-label condition, resulting in 16 children per condition, balanced for sex. Each child in the label condition was yoked to a child in the no-label condition so that both saw exactly the same animals in exactly the same configuration and order. The only difference was that one heard a label during the presentation of each test animal, and the other did not.

Procedure Toddlers were seated on their parent's lap at a table in a testing room, or at a small table on their own, with the experimenter sitting across the table from them. Each session began with a warm-up trial to familiarize them with the procedure: The experimenter demonstrated that a fish lived in an aquatic scene ("water"), and a bird lived in a tree. They were then shown additional typical fish and birds in alternating order and asked to show where each lived, until they succeeded in putting a fish in the water and a bird in the tree consecutively, or until their attention waned. Correct selections were praised, and incorrect selections were corrected.

During the testing phase, children watched as the experimenter demonstrated and explained aloud that one standard animal engaged in an activity with one prop (for example, that a cat played with a ball of yarn), and that the second standard animal of that set engaged in an activity with another prop (for example, that a dog played with a stick). Following the introduction and labeling of each standard animal, children were encouraged to imitate the activity with that same animal.

Children were then shown the two typical test animals and one of the misleading test animals from that set, one at a time and in a random order. They were asked to show the activity in which each engaged. The children in the label condition heard the experimenter use a category label to introduce each of the test animals (for example, "Look at this! Look at this dog! Can you show me what this dog plays with?"). The typical test animals were always called by labels that matched their appearance, and the misleading test animals were always called by labels that were the opposite of their appearance (i.e., if adult raters had indicated that a misleading test animal looked more like a dog than a cat, it was referred to as "this cat"). To establish baseline levels, children in the no-label condition heard the experimenter use the phrase "this one" to introduce each test animal (for example, "Look at this! Look at this one! Can you show me what this one plays with?"). Regardless of their selection, children were given neutral feedback in a positive tone ("OK!"), and the experimenter then proceeded to the next test animal or animal set.

Most children were presented with all four sets of animals; however, due to fussiness, five children were presented with three sets rather than four. The order in which the animal sets were presented was counterbalanced across pairs of children according to a Latin Square design. The prop or diorama matching a misleading test animal's perceptual appearance appeared twice on the left and twice on the right for each child, and this was counterbalanced across pairs of children.

Coding was conducted off-line, via videotape, and involved noting which of the two possible props was selected for each test animal. Two coders, blind to condition, each coded one-half of the sessions. To assess reliability, each coder also independently coded 1/4 of the other coder's sessions (selected randomly); agreement on which prop was selected was 99% (Cohen's kappa=.98).

Results and Discussion

As shown in Figure 2, children in both the label and the no-label conditions inferred that the typical test animals engaged with the props associated with their perceptual appearance significantly more frequently than would be expected by chance (50%) [label condition: $t(15)=4.01$, $p<0.01$; no-label condition: $t(15)=2.93$, $p<0.05$], and at levels not different from each other [$t(30)<1$, *n.s.*]. This is consistent with other work showing that, in the absence of labels, even 9-month-olds can make non-obvious inferences based on an object's appearance (Baldwin et al., 1993).

When a test animal was perceptually misleading, however, labeling had a dramatic effect, as Figure 2 shows. Children in the label condition made inferences based on the perceptual appearance of the misleading test animals less frequently than by chance [$t(15)=-2.93$, $p<0.05$], whereas those in the no-label condition did so at chance levels [$t(15)<1$, *n.s.*]. A 2-way mixed ANOVA on the mean percentage of perceptual inferences (label/no-label x typical/misleading) yielded a main effect of stimulus type [$F(1,30)=13.41$, $p<0.01$], and a significant interaction [$F(1,30)=6.62$, $p<0.05$]. Children in the label condition made significantly fewer inferences in line with the perceptual appearance of the misleading test animals than those in the no-label condition [$t(30)=-2.32$, $p<0.05$].

In short, children in the label condition inferred that the misleading animal engaged with the prop associated with its label, while those in the no-label condition were significantly more likely to infer that it engaged with the prop associated with its perceptual appearance. For example, when shown the misleading animal that adult subjects rated as dog-like, children who heard it referred to as "this cat" inferred that it played with the ball of yarn, while those who heard it referred to as "this one" were more likely to infer that it played with a stick.

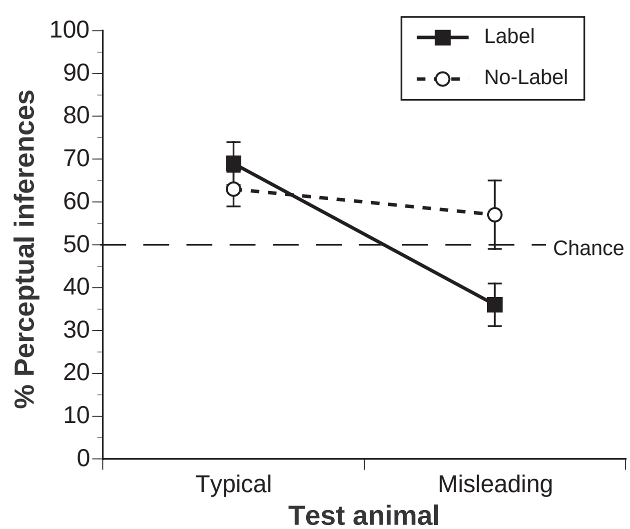

Figure 2: Mean percentage of inferences by toddlers that were consistent with perceptual appearance. Error bars show standard error.

Using a non-verbal response measure, we found that even 24-month-old children weigh the category label that a speaker applies to an animal more heavily than the animal's perceptual appearance when the two are in conflict. Although the vast majority of category labels in a typical two-year-old's productive vocabulary represent categories whose members cohere perceptually (Samuelson & Smith, 1999), these children nonetheless expect category members to share deeper similarities, and they can use language to think about objects in non-obvious—even counter-intuitive—ways. The power of a simple category label to convey non-obvious category information may well develop with age and experience with language. Certainly by 24 months, however, this process is in place, and can serve as an important mechanism for the cultural transmission of knowledge and information.

This raises interesting questions about the limits of this process, including circumstances under which children might weigh an object's perceptual appearance more heavily than its category label. In our second experiment, we explored one possibility, namely that older children might be less willing to accept a perceptually counter-intuitive category label than younger children.

Experiment 2

Experiment 2 was designed to explore whether preschoolers would be as willing to accept perceptually counter-intuitive category labels as the toddlers in Experiment 1. Presumably, older children have been exposed both to more exemplars and to a wider range of exemplars from the familiar animal categories used in our experiment. Because of this experience, they may be less willing to accept anomalous category information. For example, when they see an animal clearly possessing the perceptual features of a dog (a "misleading test animal"), they may construe it as a dog that shares properties with other dogs—regardless of what the experimenter calls it.

Method

Participants Twelve 3-year-old (M = 3 years, 3 months; range = 2-10 to 3-6) of middle-class and upper-middle-class backgrounds participated at a university-affiliated preschool. Six were girls and six were boys.

Stimuli The stimuli were the same as those used in Experiment 1.

Design Experiment 2 used a within-subject design, with the same child hearing some test animals labeled and other test animals referred to as "this one."

Procedure Children were tested individually in a quiet room at their preschool. The procedure was generally the same as that used in Experiment 1: Children watched as the experimenter demonstrated and explained aloud that one standard animal engaged with one prop, and the second standard animal of that pair engaged with another. They were then shown the three test animals one at a time and in a random order, and asked to show the activity in which each engaged. All children were shown four sets of animals.[1]

The major difference from Experiment 1 was that children heard labels for test animals from two of the sets, and did not hear labels for test animals from the other two. As in Experiment 1, children were yoked in pairs so that both saw exactly the same stimuli in exactly the same configuration and order. When one child heard test animals from a particular set labeled, his or her yoked partner did not. The order in which the sets were presented was random, as was the order of the test-baseline sets. The prop matching the misleading animal's perceptual appearance appeared equally often on the left and right for label and no-label sets.

Coding was conducted off-line, via videotape, and involved noting which of the two possible props was selected for each test animal.

Results and Discussion

As shown in Figure 3, like the toddlers in Experiment 1, preschoolers inferred that the typical test animals engaged with the props associated with their perceptual appearance significantly more frequently than would be expected by chance, regardless of whether the animals were labeled or not [label condition: $t(11)=13.40$, $p<0.01$; no-label condition: $t(11)=6.09$, $p<0.01$], and at levels not different from each other [$t(11)=1.00$, *n.s.*].

For perceptually misleading animals that were not labeled, preschoolers had them engage with the prop that matched their appearance more frequently than would be expected by chance [$t(11)=11.00$, $p<0.01$]. By contrast, when they were labeled, children had them engage with the prop that matched their appearance at chance levels [$t(11)<1$, *n.s.*]. A 2-way repeated measures ANOVA on the

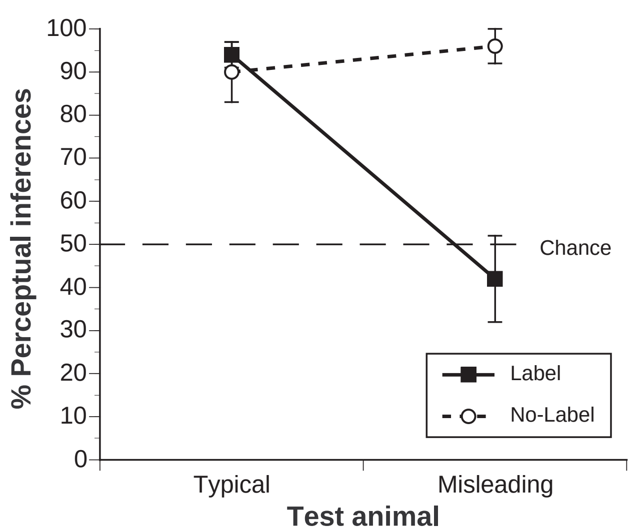

Figure 3: Mean percentage of inferences by preschoolers that were consistent with perceptual appearance. Error bars show standard error.

[1] Additionally, children were shown two sets of artifacts; those data will not be presented here.

mean percentage of perceptual inferences (label/no-label x typical/misleading) yielded a main effect of condition [$F(1,11)=33.00$, $p<0.01$], a main effect of stimulus type [$F(1,11)=7.44$, $p<0.05$], and a significant interaction [$F(1,11)=23.44$, $p<0.01$]. When they heard misleading animals labeled, children made significantly fewer inferences in line with their perceptual appearance than when they did not hear them labeled [$t(11)=-5.61$, $p<0.01$].

As in Experiment 1 with toddlers, then, labeling clearly had an effect on preschoolers' inferences about the perceptually misleading animals. However, as will be discussed below, once performance on the typical animals was equated, preschoolers appear to have been less likely to accept and use counter-intuitive category labels than toddlers.

General Discussion

These two experiments investigated children's willingness to accept perceptually counter-intuitive classifications on the basis of linguistically provided information. In Experiment 1, 24-month-olds used category labels provided by the experimenter to make non-obvious inferences about animals that were the opposite of those they would make without a label. These results converge with those of Welder and Graham (2001), who found that 16- to 21-month-olds could use a label to make a non-obvious inferences about a novel object. In that study, infants who heard two moderately dissimilar novel objects called by the same novel name inferred that they shared a non-obvious property (Study 2). When the objects were not labeled, infants did not spontaneously assume that they shared the same property (Study 1).

Our design differs from that of Welder and Graham (2001) in an important way: In particular, whereas Welder and Graham used novel objects and novel labels (except in Study 3, where they used novel objects and familiar labels), we used instead familiar objects and labels. By using familiar categories, we were able to show that despite compelling, contradictory perceptual information indicating membership in a different category, 24-month-olds nonetheless could accept and use the experimenter's label in making a non-obvious inference.

In Experiment 2, we found that preschoolers could also use labels to make inferences different from those they would make without a label. This is consistent with work by Gelman and her colleagues (Gelman & Coley, 1990; Gelman & Markman, 1986), who have found similar results with preschoolers using different procedures. However, we also found that preschoolers were less willing to accept and use an unexpected category label than the toddlers in Experiment 1. The first piece of evidence for this conclusion comes from the chance analyses, which showed that hearing a label for a misleading animal made toddlers in Experiment 1 less likely than chance to make an inference in line with the animal's perceptual appearance, whereas preschoolers in Experiment 2 were merely at chance—meaning that the older children were as likely to use the animal's perceptual appearance as the experimenter-provided label.

The second piece of evidence comes from an additional analysis of the data from the label conditions. So far, we have been considering the proportion of inferences children made in line with each test animal's perceptual appearance. Another way to consider the data from the label conditions only is in terms of the proportion of inferences children made in line with the experimenter-provided labels. These data are shown in Figure 4.

One might reasonably expect the proportion of label inferences for typical animals to be higher than the proportion of label inferences for the misleading animals, because in the case of a typical animal, there is no conflict between the label and the animal's appearance. And, in fact, as can be seen in Figure 4, for both toddlers and preschoolers, this is the case. However, whereas toddlers made only slightly fewer label inferences for misleading animals than typical ones, preschoolers were much less likely to make label inferences for misleading animals than typical ones. Indeed, a 2-way mixed ANOVA on the average percentage of label inferences (toddler/preschool x typical/misleading) yielded a main effect of stimulus type [$F(1,26)=9.67$, $p<0.01$], and a significant interaction [$F(1,26)=7.01$, $p<0.05$]. In other words, toddlers were equally likely to make an inference in line with misleading animals' labels as typical animals' labels [$t(15)<1$, *n.s.*], while preschoolers were significantly less likely to do so [$t(11)=-3.74$, $p<.01$].

Although differences in the design of these two studies make the cross-experiment comparisons somewhat tentative, this difference in children's willingness to accept anomalous category labels is extremely provocative: It suggests that 24-month-olds may be more "open-minded" about category information than 3-year-olds, perhaps because the toddlers assumed that they had not yet encountered all possible exemplars of even common

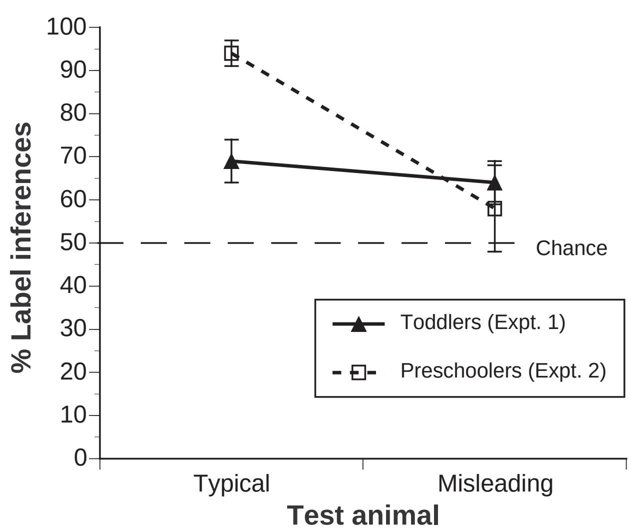

Figure 4: For toddlers and preschoolers who heard test animals labeled, mean percentage of inferences that were consistent with the labels. Error bars show standard error.

animals, whereas the preschoolers assumed they had (see Naigles, Gleitman, and Gleitman, 1992, for an analogous argument in verb learning). On-going work is considering this possibility.

Putnam (1977) has argued that a key element of adult communication is that we frequently use terms to refer to things without necessarily knowing the criteria for those terms. However, we assume that there are experts in our community who could provide the criteria and perform a test, if necessary, to fix the extension of those terms. In his words, language use requires a "division of linguistic labor." We have shown that children as young as 24 months also have something like a division of linguistic labor operating: They can accept and use what might be considered baffling category labels in order to make non-obvious inferences about animals. Using language in this way allows them to stretch the boundaries of their own spontaneously generated, often perceptually based, categories (but see Mandler & McDonough, 1996), and to take advantage of the richer and more conceptual frameworks that their cultures have evolved. A fruitful area for further investigation will be exploring the limits of this process.

Acknowledgments

We thank the children, teachers, staff, and parents who participated in this research, including those at the Arboretum Child Care Center and Bing Nursery School. Additionally, we thank A. Fernald for her assistance in subject recruitment and for providing testing space, and A. Sud for assistance in data collection. This research was supported by an NRSA predoctoral award from NIH and a Stanford Graduate Research Opportunity grant to VKJ.

References

Anglin, J. M. (1977). *Word, object, and conceptual development*. New York: Norton.

Baldwin, D. A., Markman, E. M., & Melartin, R. L. (1993). Infants' ability to draw inferences about nonobvious object properties: Evidence from exploratory play. *Child Development, 64*, 711-728.

Dale, P. S., & Fenson, L. (1996). Lexical development norms for young children. Behavior Research Methods, Instruments & Computers, *28*, 125-127.

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

Freedman, D. J., Risenhuber, M., Poggio, T., Miller, E. K. (2001). Categorical representation of visual stimuli in the primate prefrontal cortex. *Science, 291*, 312-316

Gelman, S. A., & Coley, J. D. (1990). The importance of knowing a dodo is a bird: Categories and inferences in 2-year-old children. *Developmental Psychology, 26*, 796-804.

Gelman, S. A., Hollander, M., Star, J., & Heyman, G. D. (2000). The role of language in the construction of kinds. *The Psychology of Learning and Motivation, 39*, 201-263.

Gelman, S. A., & Markman, E. M. (1986). Categories and induction in young children. *Cognition, 23*, 183-208.

Gopnik, A., & Sobel, D. M. (2000). Detecting blickets: How young children use information about novel causal powers in categorization and induction. *Child Development, 71*, 1205-1222.

Herrnstein, R. J., & de Villiers, P. A. (1980). Fish as a natural category for people and pigeons. *The Psychology of Learning and Motivation, 14*, 59-95.

Horton, M. S., & Markman, E. M. (1980). Developmental differences in the acquisition of basic and superordinate categories. *Child Development, 51*, 708-719.

Imai, M., Gentner, D., & Uchida, N. (1994). Children's theories of word meaning: The role of shape similarity in early acquisition. *Cognitive Development, 9*, 45-75.

Kemler Nelson, D. G., Russel, R., Duke, N., & Jones, K. (2000). Two-year-olds will name artifacts by their functions. *Child Development, 71*, 1271-1288.

Mandler, J. M., & McDonough, L. (1996). Drinking and driving don't mix: Inductive generalization in infancy. *Cognition, 59*, 307-335.

Markman, E. M. (1989). *Categorization and naming in children: Problems of induction*. Cambridge, MA: MIT Press.

Medin, D. L., & Ortony, A. (1989). Psychological essentialism. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning*. Cambridge: Cambridge UP.

Mervis, C. B. (1987). Child-basic object categories and early lexical development. In U. Neisser (Ed.), *Concepts and conceptual development: Ecological and intellectual factors in categorization*. Cambridge: Cambridge UP.

Mervis, C. B., & Rosch, E. (1981). Categorization of natural objects. *Annual Review of Psychology, 32*, 89-115.

Naigles, L. G., Gleitman, H., & Gleitman, L. R. (1992). Children acquire word meaning components from syntactic evidence. In E. Dromi (Ed.), *Language and cognition: A developmental perspective*. Norwood, NJ: Ablex.

Putnam, H. (1977). Meaning and reference. In S. P. Schwartz (Ed.), *Naming, necessity, and natural kinds*. Ithaca: Cornell UP.

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.

Samuelson, L. K., & Smith, L. B. (1999). Early noun vocabularies: Do ontology, category structure and syntax correspond? *Cognition, 73*, 1-33.

Smith, L. B. (1999). Children's noun learning: How general learning processes make specialized learning mechanisms. In B. MacWhinney (Ed.), *The emergence of language*. Mahwah, NJ: LEA.

Tomasello, M. (1999). *The cultural origins of human cognition*. Cambridge, MA: Harvard UP.

Welder, A. N., & Graham, S. A. (2001). The influence of shape similarity and shared labels on infants' inductive inferences about nonobvious object properties. *Child Development, 72*, 1653-1673.