Developmental Psychology
1998, Vol. 34, No. 2, 376–391

0012-1649/98/$3.00

Natural and Artifactual Kinds: Are Children Realists or Relativists About Categories?

Charles Kalish
University of Wisconsin—Madison

Research in cognitive development has highlighted important differences between conceptions of natural kinds and artifacts. One interpretation of the distinction is that natural kinds are categories one discovers, whereas artifactual kinds are invented. Four studies assessed whether children and adults saw categorization decisions as objective matters of fact or as invented conventions. Preschool-age children treated basic-level categories of animals and human-made artifacts as objective. At the superordinate level, kinds of animals were treated as more objective than were kinds of artifacts. In general, adults' judgments were similar to children's. Both children and adults have reliable and differentiated intuitions regarding category objectivity. The results from these studies are discussed in terms of their implications for structural and theory-based accounts of category naturalness.

Current theories have suggested that concepts vary in their attributes and developmental histories. One common distinction is that drawn between natural kind and artifact concepts. Researchers seem to agree that there are important differences between conceptions of natural and artifactual kinds. For example, the two types of concepts differ in the type and the number of inferences that they promote (Gelman & Coley, 1991; Markman, 1990) and in their mental representations (Keil, 1989). However, there are differing accounts of naturalness. There is agreement that concepts of living things at the basic level (e.g., *dog*) are natural kind concepts, whereas extremely artificial concepts (e.g., *nonblack thing*) are not.[1] Theories differ in their treatment of other concepts, such as those of familiar human artifacts (e.g., *chair*). The purpose of this article is to offer an additional perspective on when and how children distinguish between natural and artifactual kind concepts.

The possibility explored in the current study is that natural kinds are thought to be based on discoveries about the world, whereas artifactual kinds are viewed as having their origins in human decision. This distinction goes back to the source of much of the current theorizing about natural kinds, Locke's (1707/1961) *Essay Concerning Human Understanding*. In this work, Locke distinguished between objective discovered kinds ("substances," which have real essences) and artificial invented kinds ("mixed modes," which have only nominal essences). An artificial kind such as *gemstone* is a result of "the workmanship of the understanding." Someone decided that the features of some stones were worth noting and coined the term. What counts as a *gemstone* (and even whether the kind exists) is conventional. The kind was constructed; it is an artifact. In contrast, a natural kind such as *gold* represents our discovery that a certain sort or type of thing exists in nature. Independent of our interests or decisions, all gold is the same kind of thing. It is important to distinguish the idea that kinds may be objective or invented from the idea that instances of kinds may be naturally occurring or constructed (e.g., Gelman, 1988). For example, gold and gems are (usually) naturally occurring. The claim is that *gold* is also naturally occurring (it is a real kind in nature), whereas *gemstone* is not. Conversely, chemical elements are, nonetheless, natural kinds, even if the only samples of the elements are those created in laboratories. Mill (1872/1973) and Whewell (1840/1967) both argued that natural kinds represent real distinctions "made by nature." More recently, Kripke's (1972) theory of reference also involved a realist view of natural kinds. Thus, there is considerable precedence for interpreting the natural–artifactual distinction as relating to the status of kinds as objective or as constructed.

Philosophers are concerned with what really exists. Psychologists are interested in what people think is real (see Medin & Ortony, 1989, for a similar distinction between philosophical and psychological essentialism). Thus, the attention of philosophers will be on the kinds and categories themselves (e.g., is *dog* a natural kind?). The attention of psychologists will be on people's representations of those kinds (e.g., do people think *dog* is natural?). The suggestion explored in the empirical studies presented below is that beliefs about objectivity are important elements of people's representations of kinds. Natural kinds are those thought to be based on objective relationships. Artifactual kinds are those seen to be human-social inventions.

Given the importance that natural kinds play in cognitive development, it is important to assess children's intuitions about objectivity. The suggestion that categories may be thought to be

Portions of this research were presented at the 1995 meetings of the Society for Research in Child Development in Indianapolis, Indiana.

Preparation of this article was supported by a grant from the Graduate School of the University of Wisconsin—Madison and by a Spencer Foundation, National Academy of Education Post-Doctoral Fellowship.

I would like to thank the parents, children, and teachers at Eagle's Wing Childcare Center, the University of Wisconsin Lab School, and the Waisman Center Early Childhood Program for their participation.

Correspondence concerning this article should be addressed to Charles Kalish, University of Wisconsin—Madison, Department of Educational Psychology, 1025 West Johnson Street, Madison, Wisconsin 53706. Electronic mail may be sent to cwkalish@facstaff.wisc.edu.

[1] In this article, individuals or instances will be indicated by lowercase; concepts and categories will be indicated by italics.

objective or invented poses two question for psychologists. The first question is whether people actually make this distinction. Do children and adults have reliable and differentiated intuitions regarding category objectivity? Assuming the intuitions are robust, a second question concerns the psychological significance of this distinction. How are judgments of objectivity related to other attributes of categories? The studies reported below primarily address the first of these questions. However, this work also provides some findings that bear on the second question.

The remainder of this introduction discusses research on perceptions of objectivity and conventionality. Two accounts of category naturalness are presented that may underlie objectivity judgments. Following this discussion, four studies involving judgments of category objectivity are reported. These studies suggest that children and adults do distinguish between more and less objective categories. Further, objectivity judgments are related to other attributes that have been said to characterize natural and artifactual kinds.

Children's Judgments of Objectivity and Conventionality

Traditional perspectives on cognitive development have suggested that children progress through stages in which they have, in turn, generalized realist and relativist outlooks (Kohlberg, 1969; Piaget, 1932). In contrast, more recent studies have suggested that young children and adults share a set of distinctions between conventions and more objective arrangements. The focus of these studies has been perceptions of rules or laws. Turiel and his colleagues (Turiel, 1983, 1989; Turiel & Davidson, 1986) have argued that moral principles are judged to be objective valuations: holding independent of human decisions or norms.[2] For example, that stealing is wrong is not a matter of decision or convention about acceptable behavior; rather, it is (perceived to be) an objective truth. Similar points have been made about physical and logical laws (Komatsu & Galotti, 1986; Lockhart, Abrahams, & Osherson, 1977; Nicholls & Thorkildsen, 1988). Moral and physical laws contrast with social conventions that are understood to be particular arrangements made at the discretion of groups. A moral judgment (e.g., "Stealing is wrong.") is held to be universal and unalterable, whereas a social convention (e.g., eating with a fork or fingers) may be different in different places and can be changed (Helwig, Tisak, & Turiel, 1990; Turiel, 1983). Several studies have explored preschoolers' recognition of these differences. Children generally treat moral and physical laws as universal and unalterable but treat social conventions as more relative (Levy, Taylor, & Gelman, 1995; Smetana, 1981). Some studies find a realist bias: Across rule types, children give more objective responses than do adults (Lockhart et al., 1977; see Piaget, 1932).

Categorization decisions can be thought of as conforming to or violating categorization rules. The decision to group whales with other mammals conforms to our rules; grouping whales with fish violates these rules. As with other rules, people may see categorization norms as reflecting either relative conventions or objective facts. Natural categories involve objective rules (akin to moral laws), and artifactual categories involve invented rules (akin to conventions). Thus, the methods used in studies of children's conceptions of rules and laws are also applicable to the study of beliefs about categorization. For example, children have been asked to judge whether it might ever be acceptable to follow alternative rules: whether it could be "OK to hit" (or "OK to eat with your fingers"; e.g., Levy et al., 1995; Smetana, 1981). Accepting the alternative behavior suggests that the rule is a construction and is relative to some particular context (e.g., culture). Rejecting the alternative suggests that the rule is objective: valid or invalid, independent of human decisions. The same question can be asked about categorization rules: Are alternative ways of categorizing acceptable? For example, is it acceptable to classify whales with fish rather than with other mammals? Whether children and adults accept or reject these alternatives will depend on whether they view the categories involved as objective or conventional.

Children's Judgments of Category Objectivity

If categories may be seen as conforming to norms or rules, then the hypothesis that children are biased to treat rules as objective would also imply that they view all categories as objective. Indeed, young children have notorious difficulties understanding the constructive nature of beliefs (including, presumably, beliefs about kinds; Flavell, 1988; Wellman, 1990). Beginning about 4 years of age, children come to understand that beliefs need not match reality (Wellman, 1990). However, the idea of "match" still implies some objective standard against which truth and falsity is assessed (cf. Forguson & Gopnik, 1988). Young children may believe that all beliefs about kinds are either true or false representations of types existing in the real world. A similar prediction may be derived from young children's tendency to overascribe causality (for review, see Gelman & Kalish, 1993). Whereas adults recognize that the explanations behind some categories involve conventions, children may expect deeper reasons. Such an expectation may lead to *categorical realism*: the belief that all categories embody real objective relationships (Gelman & Kalish, 1993). Alternatively, children may recognize both objective and conventional kinds. Such an understanding would be consistent with demonstrations that children make this distinction in the realm of social rules and laws (Turiel, 1989). Thus, the first question addressed in the studies presented below is whether young children treat some categories as having an objective basis and some as having a conventional basis or whether children show a bias to treat all categories in one way or the other.

The second question addressed below is how natural and artifactual categories are to be identified. Assuming that children distinguish between objective and more conventional kinds, what accounts for this distinction? There are (at least) two major accounts of category naturalness: A category may be seen as a natural kind based on structural principles of categorization or based on theoretical beliefs. Different accounts of category

[2] Although viewing moral obligations as objective, older children (e.g., 10-year-olds) and adults nonetheless recognize that evaluations of specific acts must be made relative to the beliefs or "informational assumptions" of the actors (Wainryb, 1993). So, for example, people may accept that stealing is objectively, really, wrong but admit the possibility of disagreement as to what constitutes "stealing." Whether younger children also recognize the role of beliefs has not been demonstrated. However, children and adults do reliably state that stealing is wrong.

naturalness suggest different intuitions about objectivity. Judgments of objectivity, in turn, may inform us about which account best characterizes category naturalness.

Two Accounts of Naturalness

The broadest conception of "natural kind" rests on the claim that the human mind is constructed to form some categories and not others (Keil, 1981; Quine, 1977). Researchers have suggested that there are structural principles defining good, coherent, or natural categories (Rosch, 1973; Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976).[3] For example, Rosch argued that the human mind was tuned to form categories based on patterns of correlated attributes. There have been many different accounts of the structural determinants of category naturalness, for example, similarity, feature salience, or even familiarity (Goodman, 1955).

Categories will vary in how well they accord with structural principles of categorization. For Rosch et al. (1976), some categories are particularly well-formed (basic-level kinds, such as *cat* or *chair*). Because these categories are so "good," they may appear to be discoveries about the world. Certain groupings and relationships "cry out to be named" (Brown, 1965). For example, it seems to be an objective truth that tabbies and Siamese should be categorized together (as *cats*). Categories that are less well-structured may be seen as more arbitrary. We recognize the influence of human decision and invention in subordinate level distinctions (e.g., between *office chair* and *easy chair*). Thus, one account of category naturalness relies on structural, domain-general principles. These same principles may determine judgments of category objectivity.

Another set of accounts suggests that domain-specific beliefs make kinds seem natural or not. Natural kinds are more "theory laden" than other categories. They are organized around underlying explanatory principles rather than observable features (Gelman & Coley, 1991; Markman, 1990). Citing Boyd's notion of "causal homeostasis," Keil (1989) described natural kinds as groupings based on beliefs about shared causal interactions. Thus, kinds that play important roles in theories are natural kinds. For example, *money* is a natural kind within a theory of economics. Kinds that do not appear in a theory are artifacts. *Money* is not a natural kind within a theory of physics because there are no physical laws that apply to all or only money (see Fodor, 1975). Different theories (that focus on different types of causes and effects) will generate different natural kinds.

Theory-laden categories may be thought to be objective. Commonsense theories define our ontology: They tell us what exists (Wellman, 1990). Thus, if it is a central tenet of commonsense biology that animals are divided into species, that system of categories will seem objectively correct. Kinds that do not appear in theories will appear more arbitrary because they do not represent important similarities or distinctions in the world. Kinds of animals based on domesticated status (e.g., *pet, zoo animal*) are inventions, not discoveries. Knowledge-based accounts of naturalness suggest that people will have specific beliefs about category objectivity derived from their theories.

Summary

Natural and artifactual kind concepts seem to play important and different roles in conceptual development. One way to conceive of the difference is that natural kinds are thought to be objective, whereas artifactual kinds are seen as conventional. For example, we discover that there are *cats* but decide that there are *weekends*. Thus, it seems important to assess children's and adults' beliefs about the objectivity of categories. Although some research has explored people's understanding of objective and arbitrary rules (e.g., morals vs. social conventions; Turiel, 1983), the linkages to categories have not been made. The existing developmental literature contains some suggestions that children may be biased to treat categories as objective (and, hence, as natural kinds). The remainder of this article presents four studies of children's and adults' intuitions regarding the objectivity of categories. Two questions are addressed. First, do people (children especially) treat categorization rules as conventional or as objective, or do they recognize instances of both types? Second, which account of naturalness (structural or theory based) best accords with peoples' distinctions between objective and more conventional categories?

Study 1

Study 1 explored whether children see the bases for categorization decisions as objective facts or as invented conventions. To measure these intuitions, children were asked to evaluate alternative ways of classifying objects. Just as judgments of the propriety of alternative rules (e.g., "Hitting is OK.") have been used to demonstrate the belief that moral rules are objective and conventions are relative, judgments of alternative classifications can distinguish between objective and relative categories. As a measure of objectivity, judgments of alternative sortings were compared with judgments of alternative moral and conventional rules.

This study involved alternative sortings of animals and humanmade objects (constructs) at the basic level.[4] For example, children were asked whether it was acceptable to sort a deer together with a horse rather than with another deer. Both structural and knowledge-based accounts suggest that basic level kinds of animals should be viewed as natural and objective. Thus, these items provided a good standard for assessing the validity of the measure of objectivity. Children were predicted to reject alternative sortings of these items (It is wrong to sort a deer together with a horse rather than with another deer.). Basic-level kinds of constructs are also natural on structural principles. The qualities that determine the basic level are the same for animals and constructs. However, conceptions of animals and constructs may involve very different types or degrees of theories (Gelman & Coley, 1991; Keil, 1989). Animals, unlike constructs, may be thought to have an intrinsic and defining essence (Gelman & Coley, 1991). Thus, animals, but not constructs, may be seen as natural and objective. Because of these different predictions, comparisons of animal and construct categories will begin to address the question of the sources of objectivity judgments.

---

[3] See also Vygotsky's (1962) distinction between natural and scientific concepts.

[4] Things that people make will be referred to as *constructs*. These items are traditionally labeled *artifacts* (e.g., Gelman, 1988); however, this obscures the distinction between the instances' being human constructions and the category's being a human construction.

Table 1
Examples of Items Used in Study 1

Item type	Test item	Response options	Question asked of puppet
Animal	White-tailed deer	Mule-deer or horse	Is this one (test) the same kind of thing as this one (Response 1) or as this one (Response 2)? Which one does it go with?
Construct	Ball peen hammer	Claw-hammer or bat	(Same as for animal items)
Moral	Hitting	Naughty or nice	Is hitting another kid naughty or nice?
Conventional	Cereal	Breakfast or dinner	Is cereal a food for breakfast or dinner?
Objective control	Two things	2 pears or 1 apple	Which card has two things on it?
Arbitrary control	Preferred food	2 pears or 1 apple	Which fruit do you like better?

This study generally followed the methods used in existing studies of moral and conventional rules. However, in most paradigms, children have been asked to make judgments regarding norms that are the opposite of those that they hold (e.g., Smetana, 1981). For example, children first make their own judgment of a rule (e.g., "Is it ok to hit or not?") and then consider changing (or violating) that rule. Although these tasks have yielded impressive performance, they may place unnecessary demands on children. A child must first commit to one way of answering the question and then consider the experimenter's suggestion of a different answer. The current study removed this challenge by having participants judge a series of decisions made by a puppet. Several studies have demonstrated that children are more accurate at evaluating another's performance rather than their own (Siegal, 1991; Siegal, Waters, & Dinwiddy, 1988).

Method

Participants. Participants were 21 children recruited from a university preschool in a midsized midwestern city. Children ranged in age from 4 years 1 month to 6 years 0 months (M = 4 years 9 months). There were 10 boys and 11 girls. Participants were predominantly White and middle class.

Design. Children were asked to evaluate six types of decisions. Three instances of each type were presented. Table 1 presents examples of each type (a complete list of stimuli is given in the Appendix, Table A1). Two types included as benchmarks for comparison purposes were morals (whether a behavior is naughty or nice) and conventions (whether foods are for dinner or breakfast). Children have been found to treat social conventions, such as rules for eating, as invented and different from moral rules involving harm or rights. Morals are judged to be universally applicable (Levy et al., 1995; Smetana, 1981). Two types of items involved categorizing animals at the basic level (e.g., as a *horse* or a *deer*). Two other sets of items required categorizing constructs at the basic level (e.g., as a *boat* or a *truck*). Participants also made two "control" judgments: One involved personal preference; the second involved a logical principle. These items were included to check for possible response patterns. Actual stimuli consisted of colored line drawings presented on laminated index cards. Cards were presented for each test item as well as for each potential response. No pictures of potential responses were presented for moral items.

Procedure. Children were introduced to "Feppy" (a quasi-human puppet) and told that "Feppy comes from a place far away where they do lots of things differently than we do. Some of the things they do are wrong, but some of the things are just different." Children were asked to help the experimenter figure out when Feppy was wrong and when he was just different. Two "warm-up" judgments started the procedure. In one, Feppy incorrectly stated the child's name. In the second, Feppy chose one of two pictures as a preferred playmate. In each case, children evaluated Feppy's responses. For warm-up items, children were given corrective feedback (that Feppy was incorrect to call them the wrong name, but it was OK for him to choose either child to play with). No other feedback was provided.

Two questions elicited children's evaluations of the puppet's choices. The experimenter asked an initial question: "Is it OK for Feppy to choose (say) that or is he wrong?" (order of alternatives was randomized). For example, a child was asked whether Feppy's choice of a playmate was OK. Similarly, children also judged the acceptability of Feppy's calling them by the wrong name.

The initial response was followed-up in one of two ways, depending on whether the puppet's choice was accepted or rejected. When participants accepted the choice, they were then asked whether it "would also be OK for Feppy to choose" the other option. This question confirmed that children saw both responses as acceptable. If a child said that it was OK for Feppy to choose Playmate A (his choice) but not OK to choose Playmate B (the other), he or she was indicating that one response was correct. When participants initially rejected the puppet's choice, they were told that everyone where Feppy lives would make the same choice (say the same thing). The initial question was then re-asked. This follow-up ruled out the interpretation that Feppy had simply misspoken: Rather, he was answering according to a different norm. Thus, a positive (OK) answer to either follow-up indicated that the choice of response was judged to be relative. A negative response to either follow-up indicated the view that there was only one correct answer. This questioning strategy (initial question, then one of two follow-up questions) was used for all items in the study.

Following the warm-up items, experimental items were presented in random order, blocked with respect to type. For each type of item, the two potential responses were introduced first, and their pictures were placed in separate trays. For example, pictures of a horse and a deer were put into separate trays. Target pictures were then presented (e.g., another deer). Categorization items were not labeled (were referred to as "this one"), whereas other items were identified. Feppy was asked to make a choice (put the target card in one of the response trays; see Table 1 for wording of questions). Children then evaluated the puppet's choices.

The puppet's choices were either normative or discordant: They either agreed with or violated the culturally accepted rules that participants would endorse.[5] Feppy picked the discordant choice for two of the three

[5] Normative responses for moral, conventional, and control items were determined on the basis of the results from the self condition of Study 2. Normative responses for categorization items were assumed to be shared between the experimenter and the participants.

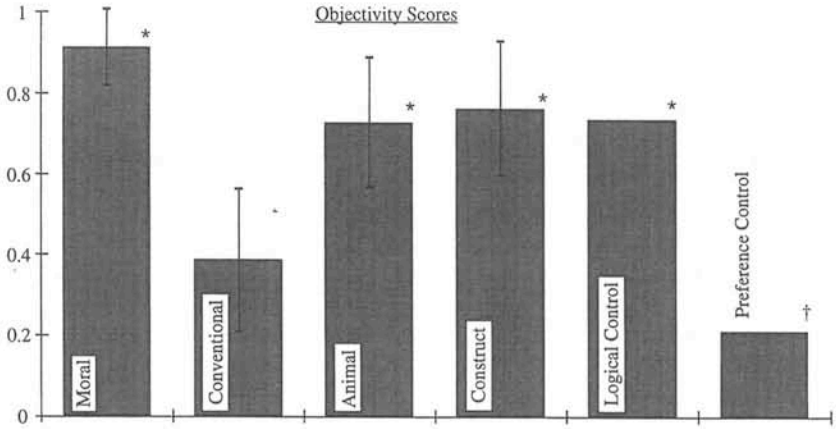

Figure 1. Mean objectivity scores (negative responses to follow-up questions) for Study 1. Higher bars indicate greater objectivity. Error bars represent 1 *SD*. $^{*}p < .05$, above chance. $^{\dagger}p < .05$, below chance.

items of a type. For example, Feppy said that salad was for breakfast, whereas cereal and corn were for dinner. The puppet's choices did not vary across children.

Scoring. In the analyses that follow, children's data will be referred to as "judgments." The puppet's answers will be referred to as "choices." The results of interest are children's judgments of the puppet's choices. Children made two judgments for each choice: They answered the initial and the follow-up questions. All responses were transformed into numerical scores (1 for *negative/rejecting* responses, 0 for *positive/accepting*). A positive answer to either follow-up question indicated that the correct choice was seen as relative to the conventions of a group. A negative answer to either follow-up indicated that the correct choice was seen as objective. The data from second judgments will be referred to as "objectivity scores." An objectivity score of 1 was assigned for each negative response to a follow-up question. Each positive response to a follow-up question was assigned an objectivity score of 0. Thus, for each item, each child received two scores: an initial response score and an objectivity score. For each child, scores were averaged across items of the same type. Thus, the data used for analyses were a mean initial response score and a mean objectivity score for each type of item for each child. Because initial responses are potentially ambiguous (see above), it is the objectivity scores that constituted the primary data for analysis. Data from initial questions are considered only when they conflict with objectivity scores. These data are reported in Table A1 in the Appendix.

Results

Benchmark items. Data from 2 participants were dropped from the study because these children judged all of Feppy's responses to be incorrect (including controls and warm-ups). Figure 1 presents the mean objectivity scores for each type of item in Study 1. The first step in the analysis was to check predictions for benchmark items. Because objectivity scores were based on responses to yes–no questions, chance performance was assumed to be 50%. As predicted, objectivity scores for morals were significantly above chance. Although objectivity scores for conventions were not significantly below chance, these items were seen as less objective than were morals, $T(17) = 148.5$, $p < .0001$.[6] Logical and preference control items also matched predictions (high and low scores, respectively). These results suggest that the procedure was effectively measuring judgments of objectivity.

Categorization items. Morals and conventions were used as benchmarks to assess whether categorization choices were viewed as objective or relative. Both animal and construct items were seen as more objective than conventions, $T(14) = 100.0$, $p < .01$; $T(15) = 110.0$, $p < .01$, respectively. In neither case were objectivity judgments different than those for morals; for animals, $T(12) = 60.5$; for constructs, $T(11) = 49.0$, both *ns*. However, in initial response data, discordant construct choices were less objective than discordant morals, $T(15) = 109$, $p < .05$, and did not differ from conventions, $T(15) = 88.5$, *ns*. Finally, there was no difference in objectivity scores for animal and construct items, $T(12) = 33.5$, *ns*, two-tailed test.

Individual patterns. The last set of analyses assessed individuals' patterns of judgments. These data are of interest for two reasons. First, they provide some indication of the responses of individual children. Rather than looking at performance of the group as a whole, we can ask whether each child showed a particular pattern. Second, it is important to evaluate the consistency of judgments across items of a type. Participants generally gave the same objectivity judgments for all instances of a type (either accepting or rejecting all alternative choices). There were three instances of each type (with two animal and two construct types). The chance probability of answering all instances of a type in the same way is .25 (binomial theorem, chance of success = .5). Thus, the chance probability of answering all instances of a type in the same way for five or six of the types is .04 (binomial theorem with chance of success = .25). Eleven of the 19 participants met the criteria for consistency.

Discussion

The results from Study 1 suggest that the method used is effective at eliciting children's judgments of objectivity. This

[6] Unless otherwise indicated, all comparisons are one-tailed Wilcoxon tests: nonparametric analogues of the *t* test. The significance levels reported for these tests have been corrected for familywise error with Holm's procedure. Holm's is a stepwise version of Dunn's procedure in which the largest test value is compared against a significance level of alpha/N (number of tests), the second largest compared against alpha/$(N - 1)$, etc. Results reported as approaching statistical significance were reliable ($p < .05$) before controlling for familywise error.

method did replicate existing findings regarding children's distinctions between morals and conventions. As predicted, participants said it was acceptable to have different social conventions. However, discrepant moral evaluations were judged to be unacceptable.

Familiar basic-level categories (animals and constructs) were treated as objective within the context of this experiment. Children generally rejected alternative sortings. Rules for categorizing animals and constructs were treated like morals rather than like conventions. Finally, objectivity judgments were fairly consistent across instances of a kind. Individual children tended to treat all the animal stimuli in the same way, for example. These findings suggest that children have relatively stable and consistent intuitions about category objectivity. Their intuitions seem to be that basic-level kinds are objective facts. Our familiar practices for sorting into these categories are seen as objectively correct and not (legitimately) violable.

Given that Study 1 explored a narrow range of categories, there are many possible interpretations of the results. On the one hand, these results may reflect a general view that each individual has one and only one way of being categorized (Markman, 1990), or some general difficulties with multiple-classification. Alternatively, responses may be based on assessments of category naturalness. Because there was no difference between animal and construct categories, one possibility is that items were treated as natural because they involved basic-level categories. As described by Rosch (Rosch et al., 1976), basic-level categories are so overdetermined by correlations between features that they appear obvious. Sorts that ignore these correlations may be unacceptable. To begin to address the sources of objectivity judgments, it is necessary to explore a wider range of stimuli.

Study 2

In Study 1, children did show reliable judgments of objectivity (across items and individuals). However, it was not clear whether children's objectivity judgments are differentiated. The primary purpose of Study 2 was to assess whether children would ever judge a categorization decision to be conventional. Thus, the stimuli of interest were arbitrary or poorly structured categories (Markman, 1990). Children were predicted to allow flexibility in sorting with these categories.

Stimuli in Study 2 were chosen to be less coherent on structural principles than were categories from Study 1. One set of stimuli were categories at a level above the basic level. The correlated attribute structure of categories is less rich at a higher (superordinate) level of generality. The superordinate categories examined in Study 2 were also chosen to be relatively unfamiliar. Thus, for animals, for example, a less familiar superordinate was tested (*feline*) rather than a more familiar category (e.g., *animal*). A second set of items involved artificial stimuli constructed to vary along two dimensions (e.g., geometric figures varying in size and shape). The attribute structure of these stimuli should allow sorting on either dimension. These items also lack the rich correlated-attribute structure that characterizes natural basic-level kinds (Rosch et al., 1976). Thus, structural principles predict that the categories used in Study 2 would be seen as less objective than those used in Study 1.

A secondary purpose of Study 2 was to explore the possibility that category domain might influence judgments of objectivity, independent of category structure. Knowledge-based accounts suggest that underlying theories determine intuitions about naturalness. One interpretation of these accounts is that kinds of animals are seen as natural (because they involve underlying essences; Atran, 1987; Gelman & Coley, 1991), whereas kinds of constructs are not. Animal categories should be judged to be more objective than are construct categories. This difference should be (at least partly) independent of the structural attributes of the categories. Domain-specific (e.g., theory-based) influences may be demonstrated by a poorly structured category being judged to be more objective than a well-structured category. Thus, it was of particular interest to choose animal categories that were weak on structural principles. Would these categories be judged to be objective despite their weak structural attributes?

Because a goal of Study 2 was to assess whether objectivity ratings were simply dependent on category structure, it was important to have some way to assess the comparative structure of categories. For this study, low intersubject agreement on category membership was taken to indicate lack of coherence on structural principles (e.g., similarity, salience, familiarity; cf. McCloskey & Glucksberg, 1978). Disagreement about what is or is not included in a category would seem to indicate that the category is relatively poorly structured. In contrast, a category for which there is high agreement on membership must be well structured in some respect. If judgments of category objectivity are based on structural principles, then we might expect that objectivity judgments will show some relationship to intersubject agreement. Categories with high agreement should be rated as more objective than categories with low agreement.[7]

Agreement on categorization decisions was assessed by having a separate group of children (in the "self" condition) categorize experimental stimuli. In addition, because the categories were unfamiliar and poorly structured, the normative choices for the items used in Study 2 were also unclear. It was not obvious which way children would prefer to sort the stimuli in Study 2. Normative categorization choices were defined as the modal membership decisions made by children in the self condition.

Method

Participants. Eighteen children recruited from a day-care center in a midsized midwestern city participated in the puppet condition. Ages ranged from 4 years 5 months to 5 years 7 months (M = 4 years 11 months). There were 8 boys and 10 girls. Eleven children participated in the self condition; there were 6 boys and 5 girls in this group (M = 5 years 0 months, range = 4 years 4 months–5 years 9 months). Three

[7] The relationship between intersubject agreement and degree of structure was not predicted to be perfect. In particular, there seems to be an important ceiling effect. There is probably some moderate level of structure (e.g., familiarity or similarity of instances) that will produce high agreement on categorization decsions. There could be very high agreement for two categories, yet one might still be more richly structured than the other. For example, both categories might be familiar, yet one also might show high similarity between instances. Because two categories might both have high agreement yet differ in degree of structure, it was not predicted that categories with similar levels of agreement would have similar objectivity scores.

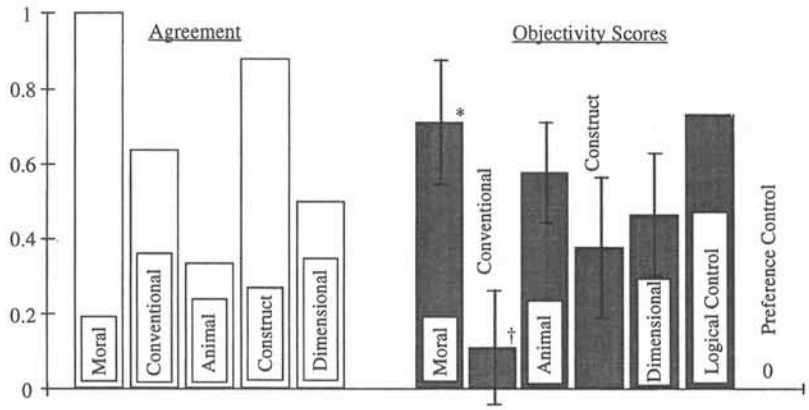

Figure 2. Mean agreement on responses from the self condition and mean objectivity scores from the puppet condition, Study 2. Higher bars indicate greater objectivity and greater agreement. Error bars represent 1 *SD*. $*p < .05$, above chance. $^{\dagger}p < .05$, below chance.

participants were eventually dropped from the puppet condition of the study because they gave the same response to all items (including controls): 2 judged all choices to be wrong, and 1 judged all to be acceptable. Children were predominantly White and from middle-class backgrounds. No child had participated in Study 1.

Design. Children were asked to make judgments about six types of items: morals, conventions, animals, constructs, and two sets of dimensional categories. A complete list of stimuli is given in the Appendix (Table A2). The moral and conventional items were the same as those used in Study 1. Animal and construct sets involved categorizing instances at a superordinate level: for example, deciding whether a lion goes with a cat or a dog. Three instances of each type were presented. Finally, there were two types of dimensional stimuli. These stimuli involved figures constructed from two binary dimensions. One set varied on shape and size (e.g., big square, small triangle). The other set varied on color and number. The logical and preference control items from Study 1 were also included. Actual stimuli consisted of colored line drawings presented on laminated index cards. Cards were presented for each test item as well as for each potential response. No pictures of potential responses were presented for moral items.

Procedure. There were two conditions. The puppet condition involved a procedure identical to that used in Study 1. In the self condition, children were simply asked to make choices about the six types of items. Participants decided whether moral items were naughty or nice, whether foods were for dinner or breakfast, and which items were of the same kind. Whereas Feppy was choosing in the puppet condition, the child was choosing in the self condition. Participants in the self condition were not asked to evaluate alternative responses. Otherwise, the procedure in the two conditions was identical (warm-up items, blocking, etc.). Data collection in the puppet condition followed completion of the self condition. Modal responses in the self condition defined the normative choices in the puppet condition.

Results

Agreement. Data from the self condition indicate the degree of intersubject agreement on responses. Figure 2 presents the mean degree of agreement for each type of item. The degree of agreement is represented as a linear transformation of the number of modal responses to range from 0 to 1 (Agreement = [mean number of modal responses − .5]/.5). Agreement did differ by type of judgment (analysis of variance, ANOVA, with type as a within-subjects variable), $F(4, 40) = 4.8$, $p < .05$. Pairwise comparisons revealed that there was significantly less agreement for animals than for constructs or moral items (Tukey's honestly significant difference, HSD; $p < .05$). No other comparisons of agreement were significant. These results suggest that animal categories were less well structured than other types of items.

Objectivity scores from the puppet condition are presented in Figure 2. The meanings of responses to initial questions depend on participants' evaluations of the puppet's choices as normative or discordant. These results are thus influenced by the degree of intersubject agreement on the normative answer. Although intersubject agreement could be assumed to be fairly high in Study 1, it is clear that there was substantial disagreement for some items in the current study. Thus, initial judgments are not considered in these analyses. These data are presented in the Appendix (see Table A2).

Benchmark items. Objectivity scores for moral, conventional, and control items showed the predicted patterns. Figure 2 presents comparisons against chance responding. Choices about moral rules were generally treated as objective, whereas choices about conventions were treated as relative (objectivity significantly below chance). Objectivity scores were higher for moral than for conventional items, $T(13) = 91$, $p < .01$. Similarly, logical and preference control items were also judged to be objective and relative, respectively.

Categorization items. Objectivity scores for animals were higher than those for conventions and did not differ from morals, $T(13) = 91$, $p < .01$; $T(10) = 38.5$, *ns*, respectively. Scores for constructs were also higher than those for conventions, $T(8) = 36$, $p < .05$. However, objectivity scores for constructs were lower than for morals, $T(11) = 60$, $p < .05$. It is important that there was also a significant difference between animal and construct items. Children treated categorization decisions for animals as more objective, $T(11) = 59$, $p < .05$, two-tailed. Thus, in these data, children distinguished between more and less objective categories.

The results for dimensional items were somewhat complex. These items were predicted to receive low objectivity scores. Dimensional categories were treated as less objective than mor-

als, $T(13) = 85$, $p < .01$, but more objective than conventions, $T(11) = 63$, $p < .05$. In some ways, it is surprising that objectivity scores for dimensional items were as high were as they were. Post hoc tests revealed a significant difference between objectivity scores for the two sets of dimensional items, $T(9) = 45$, $p < .05$, two-tailed. Objectivity scores were low for number–color items. These ratings differed from those for morals, $T(13) = 86$, $p < .05$, but not for conventions, $T(4) = 9$, *ns*. However, scores were high for size–shape items. These items differed from conventions, $T(11) = 66$, $p < .05$, but not for morals, $T(8) = 18$, *ns*. Agreement ratings for the two kinds of dimensional items also appeared to be different. There was high agreement for size–shape items (86% of responses matched the mode) but lower agreement for number–color items (64% matched the mode, minimum possible = 54%). This difference suggests that when alternative ways of sorting were salient (some children sorted by color, some by number), objectivity scores were low.

Individual patterns. A final set of analyses explored the consistency of individuals' objectivity judgments. A participant was considered to be consistent if he or she gave the same response (either accepting all choices or rejecting all choices) for at least three of the four three-item sets (animal, construct, moral, and conventional) and at least one of the two two-item sets (dimensional items). The probability of attaining this performance by chance is .04 (binomial theorem, assuming chance of success = .5). Seven of the 15 participants met the criteria for consistency.

Discussion

The results from Study 2 demonstrate that children do not treat all categorization decisions as objectively correct or incorrect. Rather, children seem to have differentiated intuitions: They view some kinds as relatively more conventional or arbitrary than others. For example, children were inflexible in judging decisions about sorting animals. There was one right way to categorize an animal, just as there was one right answer to whether an act was morally acceptable or not. In contrast, more flexibility was accepted for decisions about sorting constructs. Thus, children's judgments do not seem to reflect a general bias or response strategy to treat categories as objective.

Although children's judgments were differentiated, the determinants of objectivity judgments seem to be complex. One hypothesis is that the judgments depend on structural properties. Dimensional items did show an inverse relationship between agreement and objectivity scores. This suggests that when one dimension was very salient (always used as the basis for categorization), sorting on the other dimension was considered to be an error. Similarly, although we must be cautious comparing across studies, superordinate categories seemed to be treated as less objective than were the basic-level categories from Study 1. However, objectivity scores for superordinate items did not show a simple relationship to intersubject agreement. There was significantly lower agreement for animals than for constructs. This suggests that the perceptual cues for categorizing animal stimuli may have been weaker (and/or the categories less familiar). Nonetheless, objectivity scores were higher for animals than for constructs. Some other factors were influencing objectivity ratings. One possibility is beliefs about the content of the categories. For example, these results are consistent with beliefs that kinds of animals have essences and are more theory laden than are kinds of constructs.

Although several follow-up studies are needed to sort out the particular influences on objectivity judgments, it would also be desirable to have converging evidence that the effects from Study 2 are robust. The literature on children's distinction between morals and social conventions suggests several alternative methods of assessing objectivity. Study 3 explores children's judgments of categorization decisions with some of these alternative methods.

Study 3

Study 3 attempted to replicate the findings of Study 2 with two additional methods drawn from the literature on moral development. One measure involved judgments of the likelihood of discordant rules (e.g., a rule that it is acceptable to steal). It has been argued that people believe that discordant objective (moral) rules are less likely than discordant conventions. If asked whether there might be a place where people follow a discordant rule, children tend to answer "no" for moral rules and "yes" for conventions (e.g., Nicholls & Thorkildsen, 1988). A second measure used asked how bad it is when people follow a discordant rule. Violations of moral rules are judged to be worse than are violations of conventions (e.g., Smetana, 1981). Both of these methods seem applicable to questions about the objectivity of categories. If one way of categorizing an individual is objectively correct, then we might expect that everyone will categorize that way. Similarly, it seems to be a greater error to deny an objective truth than to deny a convention. A sort that is objectively false is worse than a sort that merely violates our social norms. Thus, judgments of the universality of acceptance and the degree of disapproval may offer converging evidence for children's views of the objectivity of categories. In addition, the method used in Study 3 asked the same children to make membership and objectivity judgments. This allowed for a more direct comparison of agreement and objectivity scores.

Method

Participants. Fifteen children recruited from a university-affiliated laboratory preschool participated in Study 3. The average age was 4 years 10 months (range = 4 years 7 months to 5 years 5 months). There were 7 boys and 8 girls in the study. Most children were White and from middle-class backgrounds. No child had participated in Studies 1 or 2.

Design and stimuli. Four types of items were included in the study. There were two instances each of moral and conventional rule judgments. There were three instance each of superordinate animal and construct categorization judgments. A complete list of items is given in the Appendix (Table A3). As in previous studies, items were presented as triads. Participants were presented with a target instance and two possible responses. Colored line drawings of each target and all possible responses were presented except in the case of moral judgments, in which only pictures of targets were used.

Procedure. The procedure was designed to mirror existing measures of objectivity (Smetana, 1981; Levy et al., 1995; Nicholls & Thorkildsen, 1988). For each target item, children were first asked to provide their choice in response to a question (e.g., "Is hitting naughty or nice?" "Is this one the same kind of thing as this one or this one?"). Two

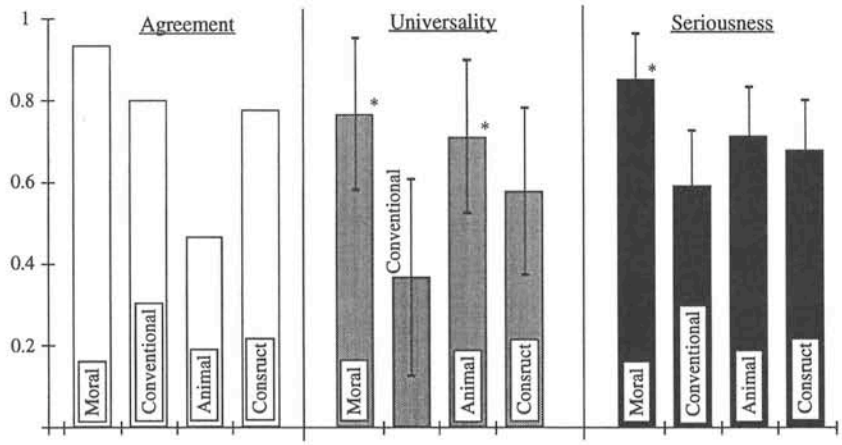

Figure 3. Mean agreement on correct choices, mean proportion of judgments that rule would not vary, mean judgments of seriousness of violation, and from Study 3. Seriousness judgments were scaled to range from 0 to 1. Higher bars indicate greater objectivity and greater agreement. Error bars represent 1 *SD*. $*p < .05$, above chance.

subsequent questions assessed objectivity. First children were asked "What if we went around to all the different countries in the world where different people lived, do you think there could be another place where people (do/choose the opposite)?" (cf. Nicholls & Thorkildsen, 1988). Participants responded "yes" or "no" to this question. A second measure asked children "What is it like when people (do or choose the opposite)? Is it great, just OK, bad, or very bad?" Responses to this question were made by indicating a position on a 4-point scale marked with pictures of smiling and frowning faces (cf. Smetana, 1981). Order of presentation of items was randomized across participants. Questioning always followed the same order: correct choice, universality, then seriousness ratings.

Results

Agreement. Figure 3 shows the mean responses for the three dependent measures used in the study. The first measure considered is children's judgments of the correct rule. Of interest is the degree of intersubject agreement. An ANOVA with item type as a within-subjects variable showed differences in levels of agreement, $F(3, 42) = 4.0$, $p < .05$. Agreement scores for animals were lower than were scores for morals, conventions, and constructs ($p < .05$, Tukey's HSD). Otherwise, agreement scores did not differ among item types.

Benchmark items. Responses to the question of whether some other people might answer differently will be referred to as universality judgments. Responses to the "What is it like?" question will be referred to as seriousness judgments.[8] Convention and moral items showed the predicted patterns. Morals were thought to be more universal than conventions, $T(9) = 42$, $p < .05$, and violations were seen as more serious, $t(15) = 3.1$, $p < .01$. Universality and seriousness judgments for morals were significantly different from chance in the directions indicating objectivity. Responses for conventions did not differ from chance. These results suggest that the method was a valid index of objectivity judgments and that moral and conventional items may be used as benchmarks for comparison with categorization items.

Categorization items. Responses to both universality and seriousness questions suggested that animal kinds were viewed as more objective than construct kinds. Animal items were rated as more universal than conventions, $T(9) = 44$, $p < .05$, but did not differ from moral items, $T(6) = 11$, *ns*. Similarly, animals and morals received the same seriousness ratings, $t(15) = 1.7$, *ns*. The difference between animals and conventions approached significance, $t(15) = 1.8$, $p = .08$. Constructs were seen as more universal than conventions, $T(8) = 35$, $p < .05$, but less than morals (this difference approached statistical significance), $T(9) = 44$, $p = .12$. These results also indicated that discordant sortings of constructs were less serious. They did not differ from conventions, $t(15) = 1.1$, *ns*, but approached significant difference from morals, $t(15) = 2.0$, $p = .09$. There was some indication that animal and construct categories differed from each other. Differences in universality ratings approached significance, $T(6) = 20$, $p = .06$, two-tailed test, though the two types of items did not differ in seriousness, $t(15) = 1.2$, *ns*. Finally, universality judgments for animals but not for constructs were greater than chance (see Figure 3).

Individual patterns. Because of the small number of items included in this study, statistical tests of individual consistency have little power. However, the results indicated that participants did respond the same way to items of the same type. For judgments of universality, it is possible to define a pattern of responding to all instances of a type the same way (i.e., either judging all to be universal or all to be relative) and doing so for all types. The probability of meeting this pattern with one or fewer deviations is .19 (binomial theorem). Eleven of the 15 participants met this criterion ($p < .00001$, second-order

[8] Seriousness judgments were scored on a 4-point scale. Values on the scale were chosen to range from 0 to 1 (0, .33, .67, 1) to enable comparisons with the universality scores that were analyzed as proportions. Similarly, with this transformation, chance performance for both measures should be 50%. Because seriousness judgments were made as ratings on a scale, they were analyzed with parametric measures (e.g., *t* tests). Universality judgments were dichotomous responses and were analyzed with nonparametric measures (e.g., Wilcoxon tests).

binomial test of 11 or more out of 15, with $p(\text{success}) = .19$). For seriousness judgments, we may ask if the child gave the same rating for all items of a type. On average, there was less than one rating difference in an individual's responses for items of a type. One rating difference would be answering *bad* for one item but *very bad* for all others. The standard deviation of a singe rating difference for moral and conventional items is .71 (e.g., rating one item as *bad* and one as *very bad*). On average, there was less than one rating difference for these types (.37 and .33, respectively). There was an average of slightly more than one rating difference for categorization items (.67). The standard deviation equivalent to one rating difference for these items is .58.

Discussion

The results from Study 3 were very similar to those of Study 2. Morals were treated as objective, conventions were treated as relative. Sorting animals into superordinate categories was treated as an objective matter of fact: as objective as moral judgments. More flexibility was allowed for sortings of constructs than for judgments of morals. These data lend some support to the finding from Study 2 that superordinate animal categories were rated as more objective than superordinate categories of constructs. Again, this result does not seem to be due to structural factors: There was less intersubject agreement for animal items than for construct items.

Although the trends from Study 3 were in the predicted directions, the results were less clear-cut with these methods. There was more variability in childrens' responses than was seen in Studies 1 and 2. This may have been because participants were first asked to make their own responses and then asked to consider the alternative reply. Children may have seen follow-up questions as challenges to their original choices (Siegal, 1991). Although the results are less clear, they do offer converging evidence that children make distinctions between more and less objective categories.

The final study included in this report investigated adults' judgments of objectivity. In considering the development of children's ideas of category naturalness, it is important to know the adult state. Study 4 investigates adults' judgments of objectivity with a paradigm similar to that used in Studies 1 and 2.

Study 4

Children's judgments were sensitive both to category content (animal or construct) and to category structure (e.g., basic or superordinate level). In particular, there was an interaction between these two factors, with content differences appearing only when category structure was weaker (only for superordinate-level kinds). It was unclear whether adults' judgments would show the same effects of content and structure. For example, adults may treat all categories as relative to cultural norms (cf. Chandler, 1988). Several studies have explored adults' flexibility in categorization (e.g., Lamberts, 1994) or beliefs about knowledge certainty as personality variables (e.g., Schommer, 1990). However, there have been few studies of adults' beliefs about whether (or which) categories are based on fact or convention.

Adults' judgments regarding the objectivity of categorization judgments have been explored in two studies. Malt (1990) asked adults how they would decide to categorize ambiguous animals and constructs. People believed that constructs could be arbitrarily assigned to categories but that there was a correct answer for categorizing animals (even if they did not know what it was). Kalish (1995) asked adults to judge whether there was a single correct answer in disputes over category membership (e.g., one person says a thing is a *chair*, another says it is not). Adults tended to accept differences of opinion for constructs but often believed there was a correct answer to disputes about categorizing animals. This pattern held for both superordinate and basic-level categories. These results suggest that adults may treat construct categories as more relativistic and conventional than animal categories.

Method

Participants. Twenty-nine students from a large midwestern university participated in the study for course credit in an educational psychology class. There were 20 women and 9 men. Participants were predominantly White.

Design. Adults were asked to evaluate (and make) all of the types of choices included in Studies 1 and 2. There were three instances of moral and conventional items, six instances of each animal and construct type, and four dimensional items. The Appendix contains a complete list of stimuli (Table A4). Importantly, the superordinate items were divided into two subtypes: familiar (*feline* and *furniture*) and unfamiliar (roughly, *invertebrate* and *food container*). An additional type of item included in this study involved judging solutions to mathematical problems. These items were included because of the possibility that college students might give relativistic responses to all items (e.g., be relativists regarding morals). For mathematical items, participants were shown an algebraic expression with one unknown and two numerals representing possible answers. Actual stimuli consisted of black-and-white scanned images of pictures from Studies 1 and 2. Data were gathered with Macintosh computers running Hypercard.

Procedure. The procedure for adults was a modified version of the tasks used in Studies 1 and 2. Participants were informed that they would be answering some questions and evaluating the answers of a third party (the "informant"). The informant was described as a person "who comes from a culture very different from ours where they do lots of things differently." The distinction between errors and differences of opinion was illustrated by one example of each (believing that the sun rotates around the earth and preferring one color to another, respectively). Thirty-seven experimental items were then presented. In each case, a participant first indicated his or her own response to the question (e.g., whether hitting was naughty or nice). The informant's response was then presented, and the participant was asked to evaluate that response as acceptable or erroneous. For two of the three instances of a type, the informant's response differed from the participant's; in one case, their responses agreed. Following each initial evaluation, participants were asked follow-up questions. Follow-up questions were the same as those used in Studies 1 and 2. If the informant's choice was rejected, participants were told that the informant was not making a mistake; rather, everyone where he lived would answer the same way. If the informant's choice was accepted, participants were asked if the alternative choice was also acceptable. Items were presented in random order, blocked with respect to type.

Results

Agreement. Agreement did differ by item type (ANOVA with type as a within-subjects variable), $F(7, 96) = 11.4$, $p < .001$. Scores fell into two groups. Moral, conventional, mathe-

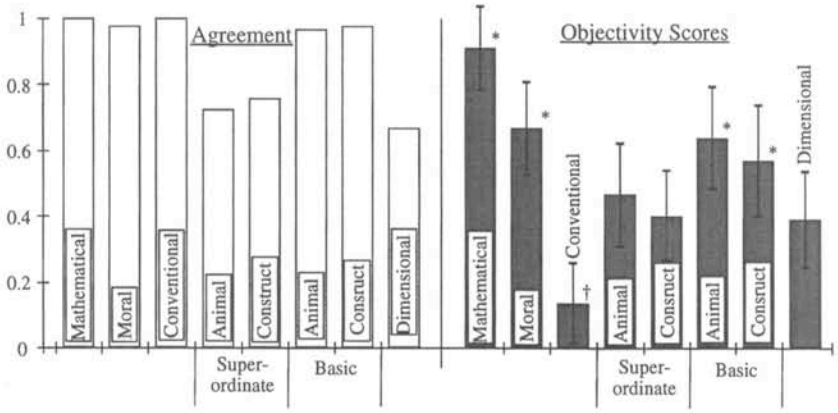

Figure 4. Mean agreement and objectivity scores from Study 4. Higher bars indicate greater objectivity and greater agreement. Error bars represent 1 *SD*. $^{*}p < .05$, above chance. $^{\dagger}p < .05$, below chance.

matical, and basic-level animal and construct items did not differ from each other. Agreement scores for these item types were significantly higher than scores for superordinate (both familiar and unfamiliar) or dimensional items. Scores for superordinate and dimensional types did not differ (all comparisons $p < .05$, Tukey's HSD).

In the current study, we may be confident that the informant's choices actually did agree or conflict with each participant's choices. Thus, participants' answers to initial questions may be viewed as more reliable indicators of objectivity than were first responses from Studies 1 and 2. Nonetheless, first responses in Study 4 must still be interpreted with caution. An initial judgment that the informant is wrong may have been based on the belief that the informant was making an idiosyncratic mistake rather than answering in accordance with a different norm. Therefore, initial judgments will be presented only when these data differed from objectivity scores. These data are reported in Table A4. The results of follow-up questions (objectivity scores) will be analyzed in detail.

Benchmark items. The three types of items included as benchmarks provided evidence that adults would treat items as both objective and conventional (relative). Mathematical and moral items were judged to be objective (only one correct answer) at greater than chance levels (see Figure 4). Judgments for these items did not differ from each other, $T(19) = 33$, *ns*. Conventional items were significantly less likely to be treated as objective than either mathematical or moral items, $T(28) = 406$, $p < .01$; $T(26) = 351$, $p < .01$, respectively. Objectivity scores for conventional items were below chance.

Categorization items. Categorization items varied along three dimensions (content, level, and familiarity). Therefore, there were several comparisons to make between different types of categorization items. The first set of comparisons addressed content differences within category levels. Comparisons across levels followed. Finally, the effects of familiarity were assessed. Dimensional items were considered last and separately from animals and constructs.

Objectivity scores for basic-level animal and construct categories did not differ from scores for moral items, $T(16) = 79$, $T(21) = 162$, *ns*, respectively. Both categories were treated as less objective than mathematical items, $T(20) = 197$, $T(20) = 206$, $p < .01$, but were seen as more objective than conventional items, $T(27) = 378$, $T(26) = 337$, $p < .01$. Finally, there was no significant difference between basic-level animal and construct categories, $T(19) = 131$, *ns*.

Superordinate-level categories of animals and constructs were treated as less objective than moral items, $T(24) = 261$, $p < .05$, $T(24) = 274$, $p < .01$, respectively, and as less objective than mathematical items, $T(23) = 276$, $T(25) = 325$, $p < .01$, respectively. These categorization items were also more objective than conventional items; animals, $T(24) = 288$; constructs, $T(24) = 289$, $p < .01$. In contrast to the results from basic-level categories, at the superordinate level, categories of animals received higher objectivity scores than did categories of constructs, $T(21) = 216$, $p < .01$. However, this difference did not appear in participants' initial responses, $T(23) = 165$, *ns*.

Analyses were conducted to compare basic- and superordinate-level categories. Superordinate categories were seen as less objective than were basic-level categories. This held true for categories of animals, $T(21) = 204$, $p < .05$, and constructs, $T(17) = 138$, $p < .05$. Thus, there is some evidence that superordinate categories were generally seen as less objective than were basic-level categories.

The effects of familiarity appeared in interaction with other aspects of categories. Familiarity moderated the effect of category level. Familiar superordinates did not differ from basic-level categories; animals, $T(24) = 213$, *ns;* constructs, $T(24) = 160$, *ns*. Unfamiliar superordinate categories were seen as less objective than basic-level categories; animals, $T(27) = 297$, $p < .05$; constructs, $T(27) = 369.5$, $p < .01$. There was also an interaction between familiarity and content. There was no significant difference between familiar and unfamiliar animal categories ($M = .51$ and .42, respectively), $T(14) = 34.5$, *ns*. However, unfamiliar constructs received lower objectivity scores than did familiar constructs ($M = .58$ and .22, respectively), $T(21) = 4.5$, $p < .01$. Unfamiliar constructs also had significantly lower objectivity scores than did unfamiliar animals, $T(15) = 19$, $p < .05$. There was no significant difference between familiar animals and constructs. These analyses suggest that unfamiliar categories were judged to be less objective than

familiar categories. However, this effect seemed to be most pronounced for constructs.

Dimensional items received intermediate objectivity scores. Scores were significantly lower for dimensional items than for morals, $T(23) = 246$, $p < .05$, or for mathematical items, $T(25) = 325$, $p < .01$, but were higher than scores for conventions, $T(21) = 215$, $p < .01$. Initial judgments of dimensional items did not differ from conventional items, $T(20) = 150$, *ns*. No significant differences were observed between the two subtypes of dimensional items, $T(14) = 22.5$, *ns*.

Individual patterns. Consistency within an item type was defined as answering five out of six items in the same way (or all the same for moral, mathematical, and conventional types that contained only three instances each). On the basis of binomial probability, the probability of answering consistently on five or more of the eight item types by chance is .02. Fourteen of the 29 participants met the criteria for consistency across item types. In the current study, both first responses and objectivity scores were reported. It is, thus, important to know how consistent participants were in making these judgments. Did first and second responses agree? For discordant choice items, responses to first questions and follow-up questions were consistent 80% of the time. Most of these nonmatching judgments (17% of total) involved judging the response to be incorrect but accepting the response once informed it represented a general group convention.

Discussion

In some ways, adults' reactions were different than children's responses. Adults' tended to see basic-level categories as more conventional than did children (compared with scores from Study 1). $U(19, 29) = 545$, $p < .05$, one-tailed Mann-Whitney test. However, responses were not significantly different for unfamiliar superordinates (the items shared in Studies 2 and 4). Adults displayed less consistency in their responses than did children. In part, this may be due to the wider range of stimuli included in the adult study. Although there were some differences, there was a general similarity in both groups' patterns of responses. Morals were considered to be very objective, conventions very relative. Judgments of categories were also similar. Both groups tended to treat well-structured (basic-level, familiar) animal and construct categories as objective. Manipulations of category attributes reduced objectivity scores. Dimensional categories were seen as less objective. Children and adults saw superordinate (and/or unfamiliar) categories as less objective than basic-level ones. For both groups, this effect was more pronounced for construct categories than for animals. When category structure was weak, kinds of animals were treated as more objective than kinds of constructs by both children and adults.

General Discussion

The results of the four studies presented above support existing claims that both children and adults make some important distinctions between natural and artifactual kinds (Gelman, 1988; Keil, 1989; Markman, 1990). In the present studies, people treated some categories as natural, objective kinds. For example, both children and adults judged that category membership at the basic level (e.g., *deer, hammer*) was a matter of fact. Alternative sortings were rejected as incorrect. In contrast, other categories were treated more like artifactual, invented kinds. Membership in unfamiliar superordinate construct categories (e.g., *food container*) was treated like a matter of opinion. In these cases, different groups could legitimately have different ways of categorizing.

Judgments of objectivity were not (simply) the results of a bias or of a general response strategy. Rather, judgments were sensitive to the categories and kinds involved. Further, objectivity judgments tended to be consistent across different instances of the same category (though the small number of items tested limits the strength of this result). The primary conclusion to draw from the results reported above is that objectivity does seem to be a recognized attribute of categories. Something about children's and adults' representations of categories leads them to accept more or less flexibility in categorization. Thus, children (and adults) seem to be neither unremitting realists nor indiscriminate relativists about categories. They recognize both natural and artifactual kinds.

Sources of Objectivity Judgments

Because participants had reliable and differentiated intuitions about the objectivity of categories, we may go on to ask about the sources of objectivity judgments. Which categories were viewed as objective and why? Conclusions here are more tentative, given the limited range of categories investigated. Two classes of influences on objectivity judgments may be distinguished. On the one hand, there are structural aspects of categories; these include familiarity, feature salience, and category level. In addition to structure, researchers have also begun to highlight the importance of theoretical beliefs about category contents (e.g., beliefs about animals vs. constructs; cf. Keil, 1989; Murphy & Medin, 1985). Though it is difficult to disentangle the various influences on judgments, the patterns of results across studies suggest that both structural and theory-based aspects of categories were affecting judgments of objectivity.

Categories strong in structural attributes were seen as objective. Basic-level kinds in Study 1 were treated this way. Categories weaker on structural attributes were judged to be less objective, more like conventions (though, see below). However, from the studies reported above, it is not possible to assess the individual contributions of particular elements of category structure. For example, category level and familiarity were confounded in the studies with children. Future studies that systematically vary such attributes as similarity, level, familiarity, and salience are needed to address the particular contribution of individual structural factors onto judgments of objectivity.

Although structural factors did influence objectivity judgments, the results suggest that these factors do not completely account for participants' responses. The assumption of these studies is that categories with higher intersubject agreement on membership were better categories on structural principles. If a category is well-structured, people should agree on classification decisions involving the category. If structural features determine objectivity judgments, then categories with lower agreement should be rated as less objective than categories with high agreement. However, in the studies reported above, animal superordinates were rated as more objective than construct su-

perordinates. This occurred despite the fact that there was less agreement for animal categories than for construct categories. The animal categories seemed to be poorer on structural measures. This reversal indicates that some other "nonstructural" factors were influencing judgments. Among the possible other influences are theoretical beliefs. For example, people may believe that in some domains (e.g., naive biology), categories are tracking the causal structure of world. In other domains (e.g., constructs), there is less causal structure, and categories are more like conventions (see Keil, 1989). Thus, the theory underlying kinds of animals would suggest that these kinds are objective features of the world. The theory (or lack of theory) underlying kinds of constructs would suggest that these kinds are invented.

Although it is premature to make definitive statements about the interaction between structural and theory-based influences on objectivity judgments, one possible interpretation consistent with the results reported above is a threshold model. There may be a level of structural coherence above which all kinds are viewed as objective. For example, the basic-level kinds investigated above might be so well structured that considerations about underlying theory do not arise or influence judgments. This is consistent with the intuition that basic-level categories "cry out to be named" (Brown, 1965). Below some threshold of structure, theoretical aspects of categories may become important. For example, kinds at a more abstract level (superordinate) or kinds that are unfamiliar are not obviously objective. They may initially appear arbitrary. However, in particular cases, the theoretical motivations and justifications underlying the grouping may override this appearance. People may believe that the kinds are objective despite their weak structural aspects. In the studies reported above, it was for less well-structured kinds (e.g., unfamiliar, superordinate categories) that theoretical beliefs seemed to play a role.

A threshold model suggests an interesting asymmetry: Theory may more readily augment the status of an otherwise weak category than degrade a well-structured kind. Theoretical beliefs may convince people that previously overlooked relationships are real and important. It may be harder to convince people that apparently real and important connections are baseless. Evidence for this asymmetry may come from the limitations of essentialism. People may believe that two animals are of a kind if they share an essence (Keil, 1989; Rips, 1989). However, this conception of animal kinds is not exclusive. People do seem to maintain that perceptual and functional features are important to animals' identities (Kalish, 1995). A system of kinds based on essences may coexist with a system based on perceptual features. Thus, although people do not seem to be complete realists or complete relativists about categories, realism may be a more dominant attitude. The philosopher John Dupré (1981) has advocated a position of "promiscuous realism": granting actual status to almost all systems of kinds. The results presented above suggest that, although they may not be promiscuous, people may be relatively unabstemious in granting reality to kinds.

Objectivity and Constraints on Concepts

Accounts of natural kinds have often been developed in the context of claims about constraints on human concepts. The argument is that human concept learners must have some way of narrowing down the infinite number of possible categorization schemes into a manageable set of plausible hypotheses (Keil, 1981). There have been debates concerning just how to characterize these constraints: as limitations, biases, assumptions, preferences, etc. One of the things researchers would like to know about a particular hypothesized constraint is its strength (cf. Keil, 1981). Does the constraint limit what is even conceivable? Does it limit which conceptions are considered to be legitimate possibilities? Or does the constraint limit which possibilities will be preferred?

Judgments of objectivity may provide a source of evidence about the strength of constraints guiding categorization. The constraints leading to basic-level categories may be fairly strong. Participants in the studies described above viewed violations of these constraints as errors. The constraints specify a correct way to categorize and an incorrect way to categorize. There is also evidence for weaker constraints: principles more akin to preferences. In the above studies, there was a relatively high degree of consensus about sorting constructs at the superordinate level. For example, people tended to judge that a bowl was the same kind of thing as a pot rather than a barrel. The high agreement indicates that there were some principles guiding these choices (people were not simply guessing). However, participants tended to be relatively accepting of alternative responses. They did not sort the bowl with the barrel, but it was acceptable for someone else to do so. It seems reasonable to describe the principles guiding these categorization decisions as preferences. Thus, there may be stronger and weaker constraints guiding categorization. Intuitions about objectivity and the acceptability of alternative categories are one way to assess these differences in strength.

Summary

One may think of categories as formed and evaluated according to rules (or constraints). Just as people recognize different types of rules governing social behavior (Turiel, 1983), they should also recognize different types of rules governing categorization. In particular, there are objective principles of categorization that underlie natural kinds and conventional principles that underlie artifactual kinds. Both children and adults made this distinction in the studies reported above. They judged some categories to be more objective or less conventional than others. Structural (e.g., familiarity, similarity) and theory-based (e.g., essentialist beliefs) attributes of categories both seemed to affect objectivity judgments. The results are consistent with a threshold model in which categories above some level of structural coherence are always judged to be objective. Below this level of coherence, theoretical beliefs come into play. Although some hypotheses were advanced, it remains for future work to carefully sort out the various influences on category objectivity. Nonetheless, these studies have demonstrated that even preschoolers have robust and consistent intuitions that some categories are more objective than others. In the context of theories of natural kinds and constraints on categories, these intuitions seem to be important sources of evidence regarding cognitive development.

References

Atran, S. (1987). Folkbiological universals as common sense. In S. Modgil & C. Modgil (Eds.), *Noam Chomsky: Consensus and controversy* (pp. 247–268). London: Falmer Press.

Brown, R. (1965). *Social psychology.* New York: Free Press.

Chandler, M. (1988). Doubt and developing theories of mind. In J. W. Astington, P. L. Harris, & D. R. Olson (Eds.), *Developing theories of mind* (pp. 387–413). New York: Cambridge University Press.

Dupré, J. (1981). Biological taxa as natural kinds. *Philosophical Review, 90,* 66–90.

Flavell, J. H. (1988). The development of children's knowledge about the mind: From cognitive connections to mental representations. In J. W. Astington, P. L. Harris, & D. R. Olson (Eds.), *Developing theories of mind* (pp. 244–267). New York: Cambridge University Press.

Fodor, J. A. (1975). *The language of thought.* Cambridge, MA: MIT Press.

Forguson, L., & Gopnik, A. (1988). The ontogeny of common sense. In J. W. Astington, P. L. Harris, & D. R. Olson (Eds.), *Developing theories of mind* (pp. 226–243). New York: Cambridge University Press.

Gelman, S. A. (1988). The development of inductions within natural kind and artifact categories. *Cognitive Psychology, 20,* 65–95.

Gelman, S. A., & Coley, J. C. (1991). Language and categorization: The acquisition of natural kind terms. In S. A. Gelman & J. P. Byrnes (Eds.), *Perspectives on language and cognition: Interrelations in development* (pp. 146–196). New York: Cambridge University Press.

Gelman, S. A., & Kalish, C. W. (1993). Categories and causality. In R. Pasnak & M. L. Howe (Eds.), *Emerging themes in cognitive development* (pp. 3–32). New York: Springer-Verlag.

Goodman, N. (1955). *Fact, fiction, and forecast.* Indianapolis, IN: Bobbs-Merrill.

Helwig, C. C., Tisak, M. S., & Turiel, E. (1990). Children's social reasoning in context: Reply to Gabennesch. *Child Development, 61,* 2068–2078.

Kalish, C. W. (1995). Graded membership in animal and artifact categories. *Memory and Cognition, 23,* 335–353.

Keil, F. C. (1981). Constraints on knowledge and cognitive development. *Psychological Review, 88,* 197–227.

Keil, F. C. (1989). *Concepts, kinds, and cognitive development.* Cambridge, MA: MIT Press.

Kohlberg, L. (1969). Stage and sequence: The cognitive-developmental approach to socialization. In D. Goslin (Ed.), *Handbook of socialization theory and research* (pp. 347–480). Chicago: Rand McNally.

Komatsu, L. K., & Galotti, K. M. (1986). Children's reasoning about social, physical and logical regularities: A look at two worlds. *Child Development, 57,* 413–420.

Kripke, S. (1972). Naming and necessity. In D. Davidson & G. Harman (Eds.), *Semantics of natural language* (pp. 253–355). Dordrecht, The Netherlands: Reidel.

Lamberts, K. (1994). Flexible tuning of similarity in exemplar-based categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 1003–1021.

Levy, G. D., Taylor, M. G., & Gelman, S. A. (1995). Traditional and evaluative aspects of flexibility in gender roles, social conventions, moral rules, and physical laws. *Child Development, 66,* 515–531.

Locke, J. (1961). *An essay concerning human understanding* (5th ed., Vol II). New York: E. P. Dutton. (Original work published 1707)

Lockhart, K. L., Abrahams, B., & Osherson, D. N. (1977). Children's understanding of uniformity in the environment. *Child Development, 48,* 1521–1531.

Malt, B. C. (1990). Features and beliefs in the mental representation of categories. *Journal of Memory and Language, 29,* 289–315.

Markman, E. M. (1990). *Categorization and naming in children.* Cambridge, MA: MIT Press.

McCloskey, M. E., & Glucksberg, S. (1978). Natural categories: Well defined or fuzzy sets? *Memory and Cognition, 6,* 462–472.

Medin, D. L., & Ortony, A. (1989). Psychological essentialism. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 179–195). New York: Cambridge University Press.

Mill, J. S. (1973). System of logic (8th ed.). In J. M. Robson (Ed.), *Collected works of John Stuart Mill* (Vols. VII & VIII). Toronto, Ontario, Canada: University of Toronto Press. (Original work published 1872)

Murphy, G. L., & Medin, D. (1985). The role of theories in conceptual coherence. *Psychological Review, 92,* 289–316.

Nicholls, J. G., & Thorkildsen, T. A. (1988). Children's distinctions among matters of intellectual convention, logic, fact, and personal preference. *Child Development, 59,* 939–949.

Piaget, J. (1932). *The moral development of the child.* New York: Harcourt Brace.

Quine, W. V. (1977). Natural kinds. In S. Schwartz (Ed.), *Naming, necessity, and natural kinds* (pp. 155–175). Ithaca, NY: Cornell University Press.

Rips, L. J. (1989). Similarity, typicality, and categorization. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 21–59). New York: Cambridge University Press.

Rosch, E. (1973). On the internal structure of perceptual and semantic categories. In T. E. Moore (Ed.), *Cognitive development and the acquisition of language* (pp. 111–144). New York: Academic Press.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology, 8,* 382–439.

Schommer, M. (1990). Effects of beliefs about the nature of knowledge on comprehension. *Journal of Educational Psychology, 82,* 498–504.

Siegal, M. (1991). *Knowing children: Experiments in conversation and cognition.* Hove, England: Erlbaum.

Siegal, M., Waters, L. J., & Dinwiddy, L. S. (1988). Misleading children: Causal attributions for inconsistency under repeated questioning. *Journal of Experimental Child Psychology, 45,* 438–456.

Smetana, J. G. (1981). Preschool children's conceptions of moral and social rules. *Child Development, 52,* 1333–1336.

Turiel, E. (1983). *The development of social knowledge: Morality and convention.* New York: Cambridge University Press.

Turiel, E. (1989). Domain-specific social judgments and domain ambiguities. *Merrill-Palmer Quarterly, 35,* 89–114.

Turiel, E., & Davidson, P. (1986). Heterogeneity, inconsistency, and asynchrony in the development of cognitive structures. In I. Levin (Ed.), *Stage and structure: Reopening the debate* (pp. 106–143). Hillsdale, NJ: Ablex.

Vygotsky, L. S. (1962). *Thought and language.* Cambridge, MA: MIT Press.

Wainryb, C. (1993). Understanding differences in moral judgments: The role of informational assumptions. *Child Development, 62,* 840–851.

Wellman, H. M. (1990). *The child's theory of mind.* Cambridge, MA: MIT Press.

Whewell, W. (1967). *The philosophy of the inductive sciences.* London: Cass. (Original work published 1840)

Appendix

Items Used in Studies 1–4

Table A1
Items Used in Study 1

Items	Options	Puppet's choice[a]	
		Normative	Discordant
Basic-level animals		.10	.79
Fish 1, Fish 2, Fish 3[b]	Frog/fish		
Deer 1, Deer 2, Deer 3[b]	Deer/horse		
Basic-level constructs		.02	.69
Hammer 1, Hammer 2, Hammer 3[b]	Bat/hammer		
Boat 1, Boat 2, Boat 3[b]	Boat/truck		
Conventions		.10	.50
Corn[b], cereal, salad	Breakfast/dinner		
Morals		.00	.95
Hitting, coloring, stealing[b]	Naughty/nice		
Controls			
Which has two items?[b]	2 pears/1 apple	.05	
Which do you like better?	2 pears/1 apple		.14

[a] Mean proportion of initial judgments that the puppet is wrong.
[b] The puppet's responses were normative for these items.

Table A2
Items Used in Study 2

Items	Options	Puppet's choice[a]	
		Normative	Discordant
Superordinate animals		.33	.60
Lion, hyena, leopard[b]	Cat/dog		
Superordinate constructs		.17	.33
Bowl, trashcan, basket[b]	Pot/barrel		
Dimensional		.23	.77
Small square, big triangle	Small triangle/big square		
2 purple[b], 3 green[b]	3 purple/2 green circles		
Conventions		.00	.13
Corn[b], cereal, salad	Breakfast/dinner		
Morals		.20	.80
Hitting, coloring, stealing[b]	Naughty/nice		
Controls			.73
Which has two items?	2 pears/1 apple		
Which do you like better?[b]	2 pears/1 apple	.00	

[a] Mean proportion of initial judgments that the puppet is wrong.
[b] The puppet's responses were normative for these items.

Table A3
Items Used in Study 3

Items	Options
Animals	
Frog	Fish/crab
Rabbit	Mouse/turkey
Lion	Cat/dog
Constructs	
Train	Truck/boat
Scissors	Ruler/knife
Broom	Brush/basket
Morals	
Hitting	Mean/OK
Sharing	Naughty/nice
Conventions	
Cereal	Breakfast/dinner
Corn	Breakfast/dinner

Table A4
Items Used in Study 4

		Informant's choice[a]	
Items	Options	Agrees with participant's	Disagrees with participant's
Basic-level animals		.03	.74
Fish 1, Fish 2, Fish 3[b]	Frog/fish		
Deer 1, Deer 2, Deer 3[b]	Deer/horse		
Basic-level constructs		.05	.64
Hammer 1, Hammer 2, Hammer 3[b]	Bat/hammer		
Boat 1, Boat 2, Boat 3[b]	Boat/truck		
Superordinate animals		.03	.60
Lion, hyena, leopard[b]	Cat/dog		
Lobster, hummingbird, crab[b]	Bee/turtle		
Superordinate constructs		.05	.56
Bowl, trashcan, basket[b]	Pot/barrel		
Piano, television, harp[b]	Guitar/desk		
Dimensional		.03	.36
Small square, big triangle	Small triangle/big square		
2 gray[b], 3 white[b]	3 gray/2 white circles		
Morals		.10	.81
Hitting, coloring, stealing[b]	Naughty/nice		
Conventions		.03	.24
Corn[b], cereal, salad	Breakfast/dinner		
Mathematical		.03	.97
$5 + 7 = 12$ or 14[b]; $2 \times 3 = 8$ or 6; $4 - 8 = -4$ or 0			

[a] Mean proportion of initial judgments that the informant is wrong.
[b] The informants' responses matched participants' responses.

Received October 3, 1996
Revision received September 29, 1997
Accepted September 29, 1997 ■