



Published in final edited form as:

Cogn Psychol. 2013 May ; 66(3): 327–353. doi:10.1016/j.cogpsych.2013.02.001.

Conceptual influences on category-based induction

Susan A. Gelman and

Department of Psychology, University of Michigan

Natalie S. Davidson

Department of Psychology, University of Michigan

Abstract

One important function of categories is to permit rich inductive inferences. Prior work shows that children use category labels to guide their inductive inferences. However, there are competing theories to explain this phenomenon, differing in the roles attributed to conceptual information versus perceptual similarity. Seven experiments with 4- to 5-year-old children and adults ($N = 344$) test these theories by teaching categories for which category membership and perceptual similarity are in conflict, and varying the conceptual basis of the novel categories. Results indicate that for non-natural kind categories that have little conceptual coherence, children make inferences based on perceptual similarity, whereas adults make inferences based on category membership. In contrast, for basic- and ontological-level categories that have a principled conceptual basis, children and adults alike make use of category membership more than perceptual similarity as the basis of their inferences. These findings provide evidence in favor of the role of conceptual information in preschoolers' inferences, and further demonstrate that labeled categories are not all equivalent; they differ in their inductive potential.

One central function of categories is to permit rich inductive inferences (e.g., apples grow on trees and have seeds; birds lay eggs and have hollow bones). People readily extend observations with a limited number of instances to other category members (Nisbett, Krantz, Jepson, & Kunda, 1983) or even to the category as a whole (Leslie, 2008). Thus, category-based induction is a powerful tool for extending knowledge and constructing theories in a variety of content domains (Murphy, 2002; Osherson, Smith, Wilkie, López, & Shafir, 1990; Rips, 1975).

An important question in the study of category-based induction is the role of language. Although induction does not require language (Baldwin, Markman, & Melartin, 1993), supplying a category label for a group of items has been shown to foster inductive inferences. For example, the presence of a shared label helps infants and young children to categorize dissimilar things and treat them as alike (Dewar & Xu, 2009; Waxman, 2004), and adults show similar effects (Lupyan, Rakison, & McClelland, 2007). Furthermore, preschool children as well as adults are more likely to infer that dissimilar members of a category share properties when they are labeled than when they are not (e.g., a blackbird shares more properties with a flamingo than a bat when the items are labeled as “bird”,

© 2013 Elsevier Inc. All rights reserved.

Address for correspondence: Dr. Susan Gelman, University of Michigan, 530 Church St., Department of Psychology, Ann Arbor, MI 48109-1043; gelman@umich.edu. Telephone: 734.764.0268, FAX: 734.615.0573.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

“bird”, and “bat”; Gelman & Markman, 1986, 1987). Labels guide children’s inferences about novel artifacts as early as 13 months of age (Graham, Kilbreath, & Welder, 2004); for example after learning that a novel object makes a rattling sound, infants were more likely to expect that another object would make the same sound if it was given the same label.

Two classes of accounts have been offered to explain this result. They can be roughly characterized as being based on “perceptual similarity” versus “conceptual” information (though, as will be noted, neither account strictly excludes the other). On the *perceptual similarity* account, children’s inferences are driven by perceptual similarity, so that children draw many inferences from one item to a visually similar item, and few inferences from an item to a visually dissimilar item. To the extent that labels affect inferences, they do so only by contributing to the perceived perceptual similarity among items (Sloutsky, Kloos, & Fisher, 2007). Thus, on this account, items that receive identical labels (e.g., “dog” applied to both a dachshund and a corgi) will be viewed as more similar to one another in the context in which the words are spoken, because of the shared auditory feature. The label becomes one more perceptual feature associated with the item, with the consequence that shared labels contribute to shared similarity. Although this view does not explicitly account for the role of conceptual information, such information may be required to determine which labels are relevant (e.g., intentional vs. unintentional naming, Akhtar & Tomasello, 2000; naming a real object vs. a representation of that object, Gelman & Ebeling, 1998; Massey & Gelman, 1988).

In contrast, on the *conceptual* account, labels denote category membership, and category information drives induction (Gelman, 2003). For example, hearing that an animal is a lizard conveys that it belongs to the same category as other lizards, and thus that it shares properties with them. The conceptual account also has a corollary, which is that categories are variable and thus differ in their inductive potential (Waxman & Gelman, 2009). Whereas some categories are richly structured “natural kinds” that capture clusters of features and provide a firm basis for inductive inferences (e.g., “dogs”), other categories are arbitrary, share only one or a few properties, and provide a much weaker basis for inductive inferences (e.g., “spotted things”, a category that includes Dalmatians, ladybugs, spotted stones, polka-dotted shirts, etc.). The conceptual account does not exclude the relevance of perceptual similarity. To the contrary, the conceptual account assumes that perceptual cues are important for identifying features, and that perceptual similarity provides a useful cue to category membership. Typically, perceptual similarity and category membership are highly correlated (e.g., a certain body shape typically implies that an animal is a snake; Diesendruck & Bloom, 2003; Gelman & Medin, 1993). However, according to the conceptual view, perceptual similarity is an imperfect cue to category membership (legless lizards look more like snakes than lizards), and perceptual similarity alone is insufficient to account for children’s inductive inferences (Waxman & Gelman, 2009).

In a paper designed to test these competing theories, Sloutsky et al. (2007) (hereafter referred to as SKF) taught 4- and 5-year-old children two novel categories (ziblets and flurps) that were specially designed so that category membership and perceptual similarity could be placed in direct conflict, thus permitting a test of which factor was more influential in children’s inductive inferences. As shown in Figure 1a, testing triads were constructed such that perceptual similarity directly conflicted with category membership. For example, in the set provided in Figure 1a, the target item and the item on the bottom left are both flurps, whereas the target item and the item on the bottom right both have more similar appearance. More generally, ziblets were just as similar to flurps as they were to other ziblets; likewise, flurps were as similar to ziblets as they were to other flurps.

Furthermore, in SKF, although the categories were labeled during a training phase, there was also a clear rule for identifying an animal's category membership, so that during the testing phase, children's inductive inferences could be examined without the label being present. Specifically, the two categories could be distinguished based on the ratio of fingers to buttons: ziblets had more fingers on one hand than buttons, whereas flurps had more buttons on one hand than fingers. This is important, as the perceptual similarity-based approach would predict that the mere presence of a label during testing would be sufficient to evoke category-based inferences. Thus, the design of the experiment permits a test of whether children use category membership or appearances as a guide.

Using this approach, SKF found that young children were far more likely to use perceptual similarity than category membership to guide their inductive inferences. Thus, if a round, green flurp was said to have a big brain, then 4- and 5-year-old children generalized this fact to a round, green ziblet rather than to a differently shaped, yellow flurp. Children's failure to use category membership was striking and pervasive. In a follow-up study in which perceptual similarity was uninformative regarding category membership (neither positively nor negatively correlated with it; i.e., appearances did not pose a competing factor), children still failed to use category membership as a guide to their inductive inferences, instead answering randomly. SKF conclude (p. 183): "The main finding of these two experiments is that young children induced hidden properties on the basis of appearance similarity rather than on the basis of shared membership in a natural-kind category. ... [Early] induction is similarity based and not category based."

Although children made appearance-based inferences in this task, it is premature to conclude that early induction is based on perceptual similarity across the board. Gelman and Waxman (2007) noted several aspects of SKF that might limit or reduce children's reliance on category information and conversely encourage the use of perceptual features. Their central argument was that the novel categories created for these experiments were not natural kinds, although they were characterized as such in the article (SKF, pp. 180, 183, 184). Natural kinds (such as tigers, oak trees, or gold) are believed to reflect basic divisions in nature, to be non-arbitrary, and to capture a rich cluster of non-obvious properties, both known and unknown (Keil, 1989; Schwartz, 1977). In contrast, ziblets and flurps were defined by a single, arbitrary, non-biologically-plausible property: fingers-to-buttons ratio. That the ziblet-flurp distinction rests on a non-biological property ("buttons") suggests that the distinction reflects superficial features (i.e., clothing) rather than inherent or natural qualities (see Rhodes & Gelman, 2008, Study 4, for evidence that children reason differently about clothing-based categories than natural kinds). Ziblets and flurps also were characterized as differing in personality (ziblets are nice and friendly; flurps are wild and dangerous). Importantly, however, personality can be an individual difference within a kind rather than relevant to distinguishing kinds (e.g., for many natural kinds, including humans, dogs, and cats, some individuals within a kind are friendly whereas others are dangerous). A related concern was that the category distinction could be construed to be subordinate-level rather than basic-level, because none of the signature properties known to characterize basic-level kinds—such as diet, habitat, means of locomotion, or vocalizations (Shipley, 1993)—were present. Prior research has documented that preschool children do not spontaneously treat subordinate-level categories as distinct kinds (Waxman, Lynch, Casey, & Baer, 1997).ⁱ

The conceptual account predicts that natural kinds will serve as a guide for children's inferences, but that categories that are not natural kinds will not serve that purpose. Any grouping of instances could constitute a category, but only some categories guide induction. For example, contrast *tiger* (a natural kind) with *striped things* (not a natural kind). The category *tiger* permits rich inferences, but the category *striped things* (including tigers, barber poles, and candy canes) does not. For these reasons, Gelman and Waxman (2007)

argued that for the ziblets and flurps in SKF, appearances were at least as strong a guide to natural-kind membership as were the labels. Thus, the primary concern is that the categories selected for study were not natural kinds and thus did not provide a test of the conceptual account.

More recently Badger and Shapiro (2012) conducted two inductive inference studies that were modeled on SKF's original work but addressed some methodological limitations. For example, in SKF, the distinguishing feature that allowed one to determine whether an instance was a ziblet or flurp was very effortful to identify, thereby possibly discouraging use of the category when not explicitly directed to do so. This may be analogous to the "production deficiency" that children exhibit when given a memorization task, in which they often avoid using mnemonic strategies unless given explicit instructions (Moely, Olson, Halwes, & Flavell, 1969). Badger and Shapiro's categories had a more readily accessible identifying feature (shape of head rather than finger:button ratio), and the categorical distinction was more biologically grounded (habitat and behavior). Results with children 3–9 years of age indicated a developmental increase in the use of the target categories (sandbugs vs. rockbugs), and thus were interpreted as supporting the perceptual similarity account. However, the implications of these studies are unsettled. First, even the youngest children (3- to 4-year-olds) made use of the category information more than one would expect on the basis of similarity ratings of the stimuli, and by 4–5 years of age, children selected the category choice as often as the perceptual choice. Thus, the data suggest that children in all age groups valued the categories as inductively useful, despite overwhelming perceptual evidence in opposition to this choice. Second, the primary distinction (sandbugs vs. rockbugs) was a subordinate-level distinction of two kinds of bugs. If, as noted earlier, children fail to treat subordinate-level distinctions as natural kinds, instead using basic-level categories as the basis for inductive inferences (Waxman et al., 1997), then either choice would be a plausible categorical basis for induction. Finally, in one of the studies, the "perceptual" choice was in fact itself a cross-cutting category choice (namely, age grouping: baby vs. adult), which itself has powerful inductive potential (Taylor & Gelman, 1993).

Although Sloutsky et al. (2007) and Badger and Shapiro (2012) concluded that children make use of perceptual similarity, not category membership, in their inductive inferences, a major limitation in this previous research is that the conceptual foundation for the test categories was not systematically varied. As noted earlier, the conceptual account proposes that children are flexible in their inductive inferences, using perceptual similarity when categories have a weak conceptual basis, but using category membership for categories with a stronger conceptual basis. This prediction is supported by prior research demonstrating that by preschool age, children are flexible in their inductive inferences. For example, Gelman (1988) found that preschoolers make category-based inferences concerning inherent, non-obvious properties but not transient properties. Similarly, Davidson and Gelman (1990) found that children were flexible and selective in their use of novel labels as a basis for inductive inferences. When labels were completely crossed with perceptual similarity, children did not make use of the label as a basis for induction, but when labels had some perceptual support, children did so. Likewise, Booth (2012) found that "novel

¹It is unclear why these seemingly arbitrary groupings were characterized as "natural kinds". Perhaps it was assumed that providing labels (ziblet vs. flurp) transforms a category into a natural kind. This would be an invalid assumption, however, because labels apply not only to natural kinds (rabbits, gold) but also situation-restricted categories (passengers), temporary groupings (students), phase sortals (droplets), relationship-based entities (pets), etc. A second possible rationale for treating ziblets and flurps as natural kinds was that, in one experiment (SKF, Experiment 1, second control, p. 183), children heard why the finger-button ratio was high for ziblets: "Ziblets have more fingers than buttons because they catch food with their fingers not with their buttons. The chemical Zeeken makes their fingers sticky so they can catch food with their fingers. Ziblets don't need their buttons for anything, so they don't have so many buttons." However, this would seem more relevant to raw numbers of fingers and buttons (and possibly finger length), rather than finger-button ratio. Furthermore, nothing was said regarding how flurps eat, or how they use their fingers and buttons. Indeed, children never learned that flurps don't have Zeeken, and thus the property could not be assumed to differentiate the categories.

words do indeed support inductive inference, but only when they are known by children to reference causally rich categories.”

Present experiments

The present experiments were designed to examine whether young children’s patterns of inductive inferences are sensitive to the conceptual structure of the test categories. As noted previously, the conceptual account predicts that categories that are richly structured natural kinds will support children’s inductive inferences, but that categories that have an arbitrary basis will not support children’s inductive inferences. This is in contrast to the perceptual similarity view, which posits that all labels are equivalent in providing an added perceptual feature, so that category type should not matter.

There are at least four factors that may signal that a category has a relatively stronger conceptual basis (i.e., is more “natural kind”-like than arbitrary). First, the level at which categories differ from one another will affect the conceptual distinctiveness of the categories (Rosch, 1976). Thus, categories that differ at an ontological level (e.g., animal vs. artifact) will be conceptually highly distinct, whereas categories that differ at a subordinate level (e.g., two kinds of dogs) will not. Second, the richness of shared features within a category is a key signature of natural kind categories (Mill, 1843; Gelman & Markman, 1986; Schwartz, 1979), with certain features being particularly predictive of biological kind membership. For example, members of an animate natural kind tend to share diet, habitat, vocalization, form of locomotion, form of reproduction, etc. (Shipley, 1993). Thus, categories that differ in several of these signature dimensions will be conceptually rich, whereas categories that are not known to differ on any of these dimensions will be less so. Third, the nature of the distinguishing feature can indicate whether categories are likely to be conceptually distinct. For example, categories that differ from one another according to inherent, functional features are more distinct than categories that differ from one another according to temporary or arbitrary features such as number of buttons or eyelashes. Finally, emphasizing how two categories contrast with one another can highlight the conceptual basis of the distinction (Namy & Clepper, 2010).

In order to examine the role of labels in children’s inductive inferences, we designed a series of experiments in which children learn categories that vary in their conceptual distinctiveness. We hypothesized that when children are presented with categories that are conceptually distinct, they will use the label as a cue to category membership and as the basis for inductive inferences. However, we also hypothesized that when children are presented with categories that are not clearly conceptually distinct, they will fail to use the label as a cue to category membership. We also examined how children’s performance compares to that of adults. Based on the work of Badger and Shapiro (2012), we predicted that adults would infer that distinct labels refer to distinct categories, regardless of whether they received overt evidence that the categories are conceptually distinct.

We conducted seven experiments, all based on SKF’s original design, but varying aspects of the materials, task, and procedure (see Table 1). Experiment 1 was a direct replication of SKF with two age groups, children (4 and 5 years of age) and adults. A replication was necessary because our experimental materials differed slightly from those used by SKF. This also allowed us to include an assessment of adult performance (which had not been examined previously). Experiment 2 was modeled directly on Experiment 1, but the categories were modified such that ziblets and flurps were from ontologically distinct domains (animals and machines, respectively). Furthermore, the test properties in Experiment 2 were familiar properties that are linked to animals or machines (e.g., “has a heart inside”; “has a wire inside”). Thus, Experiment 2 tested the relative importance of

perceptual similarity versus category information when considering properties that are embedded in children's established factual knowledge of animals and artifacts. Experiment 3 was identical to Experiment 2, except that the test properties were all wholly novel (e.g., "has blickets inside"). Experiment 4 was identical to Experiment 3, except that ziblets and flurps were from the same domain and contrasted at the basic level (i.e., two kinds of animals) rather than from distinct domains. Finally, Experiments 5, 6, and 7 were designed to control for aspects of the stimuli and procedure that were modified in Experiments 2–4. Each of Experiments 2–6 also included a control version in which participants received the induction task first, before categorization, to independently assess the influence of categorization information on performance.

Experiment 1: Non-Kind Categories, Novel Properties (Replication of Sloutsky et al., 2007)

Experiment 1 has three purposes: (1) to replicate SKF, but with stimuli that were slightly modified in order to fit the requirements of the design, (2) to include adult participants, and (3) to permit a direct statistical test comparing children to adults.

Method

Participants—Participants were 16 children (M age = 4.56, SD =0.59; 5 boys, 11 girls) and 24 adults (M age = 19.25, SD =0.99; 10 men, 14 women). Five additional children were tested but excluded from the sample because they did not reach criterion of 75% correct on the initial categorization Task. Children were predominantly middle-class and white. Adults were undergraduate students in an Introductory Psychology class. All participants were from a midsize city in the Midwestern United States.

Materials—Materials included drawings of: a boy (Fritz), a pet store, 2 individual creatures (both ziblets) used during the category training phase of the experiment, 32 individual creatures used in the category learning and initial categorization phases, and 36 individual creatures used in the induction and final categorization tasks. The creatures used during category training were created for this experiment but modeled directly on others used by SKF; the creatures in the category learning and initial categorization phases included 16 ziblets (8 with appearance A1, 8 with appearance A2) and 16 flurps (8 with appearance A1, 8 with appearance A2), selected from those used by SKFⁱⁱ (see Figure 2). However, each participant saw only 16 of the 32 creatures (4 ziblets and 4 flurps during category learning; 4 ziblets and 4 flurps during initial categorization; within each set of 4, 2 had appearance A1 and 2 had appearance A2). Assignment of creature to the category learning task or the initial categorization task was completely counterbalanced across participants, using 4 sets of assignments.

All of the induction sets had the structure represented in Figure 1a, where one choice matched the target on category membership but not appearance, whereas the other choice matched the target on appearance but not category membership. For the induction task, we included 12 of the 16 sets that SKF had originally usedⁱⁱⁱ: 6 sets included ziblets as the target creature and 6 sets included flurps as the target creature; the targets all had appearance A1. The side on which the test items appeared was counterbalanced across trials. The properties used in the induction task included all 8 of those used by SKF, plus an additional 4 that we created so as not to introduce any repeated properties, as SKF had done (see Table 2).

ⁱⁱWe thank SKF for sharing their original stimuli and script with us.

ⁱⁱⁱOne of the items was slightly modified from SKF's original set. In the original set, one of the flurp test items displayed an equal number of fingers as buttons; we modified the image so that there were more buttons than fingers.

Procedure—Participants were tested individually. Children were in a quiet room, either at their preschool or in an on-campus laboratory. Adults were in a quiet room in the on-campus lab. Materials were presented on a computer using PowerPoint software. The experimental procedure had five phases: category training, category learning, initial categorization, induction, and final categorization (see Table 3 and Appendix). The wording of all phases was identical to that of SKF's original wording. During *category training*, the researcher introduced the distinction between ziblets and flurps, explained the categorization rule (ziblets have more fingers than buttons, whereas flurps do not), and illustrated the rule with two examples of ziblets. The *category learning* task was introduced by mentioning that participants would see some animals from “the magical pet store”. In the task itself, participants saw 8 creatures (4 ziblets and 4 flurps), one at a time and in randomized order, and were asked to categorize each as a ziblet or a flurp. Responses could include either the words “ziblet” or “flurp”, or mention of the relevant features (e.g., “more fingers than buttons”). After each response during this phase, the participant received direct feedback regarding the basis of the categorization (e.g., “Yes, it’s a ziblet. It had more fingers than buttons”; “Yes, it’s a flurp. It did not have more fingers than buttons”), including corrections when necessary. *Initial categorization* was identical to category learning, except that participants saw a different set of 4 ziblets and 4 flurps, and received no feedback. Participants were included only if they were correct on at least 75% of the initial categorization trials (6 out of 8). The *induction task* was introduced by telling the participant that the pet store owner had a few questions for those who wanted to buy a pet. The task itself consisted of 12 trials. On each trial, participants saw a triad as described in the Materials section (above) and Figure 1a, learned a new fact regarding the target creature, and were asked which test creature also possessed the property (e.g., “This creature has two stomachs. Which of these creatures has two stomachs, too?”). No feedback was provided. The *final categorization* task was identical to the initial categorization task, except that the items included 4 ziblets and 4 flurps that were a subset of the test items in the induction task.

Results

First we tested to ensure that participants learned and remembered the relevant categories of “ziblets” and “flurps”. Participants who passed the categorization criterion were highly accurate in the initial categorization task, with *M*s (out of 8) of 7.44 (children) and 7.96 (adults), both greater than chance, $ps < .001$. Importantly, performance remained highly accurate in the final categorization task, demonstrating that participants remembered the categorization rule throughout the experiment; *M*s of 5.94 (children) and 8.00 (adults), both greater than chance, $ps < .01$.

The primary analyses focused on responses in the induction task. Each participant received a score indicating the number of trials (out of 12) on which they selected the categorical response (see Table 4). We conducted a simple independent-sample *t*-test, comparing the scores of children versus adults. As predicted, adults provided more category-based responses ($M=8.83$) than children ($M=3.31$), $t(38) = 5.52$, $p < .001$. Whereas adults selected the category-based response significantly above chance (of 6.0), $p = .001$, children selected the category-based response significantly below chance, $p < .001$.

Discussion

In Experiment 1, on an inductive inference task, children were more likely to use perceptual similarity than category membership as the basis for their inferences, despite having learned the classification rule for the novel categories and despite successfully identifying category membership on the basis of the rule, both before and after the inductive inference task. These findings replicate earlier work with children, pitting category membership against perceptual similarity when the conceptual basis of the category distinction is minimal, and

the distinguishing characteristic is difficult to assess (Sloutsky, Kloos, & Fisher, 2007). Furthermore, Experiment 1 further finds that adults were more likely to use category membership than children, and more likely to use category membership than chance.

Experiment 1 provides a foundation for the studies that follow. In Experiments 2–6, we started with the same basic structure of the design of Experiment 1 but varied the content of the categories and property inferences to determine conditions under which children and adults do and don't make use of category membership in their inductive inferences.

Experiment 2: Ontologically Distinct Categories, Familiar Test Properties

Experiment 2 made use of the same general procedure as Experiment 1 (category training, category learning, initial categorization, induction, and final categorization), using the pictures from Experiment 1 as a foundation (see below), and the same category labels as in Experiment 1 (ziblets and flurps). However, Experiment 2 made four modifications designed to enhance the conceptual distinction between the two novel categories:

- The *categories* were from two ontologically distinct domains (ziblets were a type of animal, flurps were a type of artifact). In contrast, in Experiment 1, ziblets and flurps were two types of pets and did not differ in any of the properties typically associated with distinct kinds, such as diet, habitat, means of locomotion, or vocalization (see Shipley, 1993, for discussion of these kind-relevant property dimensions). Because ziblets and flurps were no longer types of pets, we removed all mention of pets, the pet store, and the pet store owner (see Appendix).
- The *distinguishing features* were biologically meaningful and readily apparent. Thus, ziblets have eyes to help them see; flurps have bolts that keep them together. The eyes and bolts, though small in size, can be discerned at a glance. In contrast, the distinguishing features used in Experiment 1 did not correspond to a meaningful biological attribute (ratio of fingers to buttons), and required effort to detect (counting fingers, counting buttons, and making a comparison).
- The *associated features* were stable and inherent, and represented principled, kind-based distinctions (i.e., concerning diet and habitat; as noted earlier, diet and habitat are signatures of distinct basic-level categories). Thus, ziblets live in trees and eat grapes; flurps come from factories and use electricity. In contrast, the associated features used in Experiment 1 were potentially variable both within a category and within an individual (i.e., concerning personality characteristics: nice vs. mean).
- The *category training* and *categorization tasks* emphasized the contrast between the two categories (e.g., both ziblets and flurps were provided and contrasted during category training), and participants were reminded of the conceptual basis of the category distinction, by asking them to sort instances into appropriate habitats (tree vs. factory). In contrast, the category training task in Experiment 1 focused exclusively on ziblets, and the categorization task used in Experiment 1 did not include reminders of the conceptual basis.

In addition to these major changes, minor modifications to the wording were introduced in order to make the task more familiar and child-friendly, and to reduce information-processing demands. For example, the wording was streamlined, the character was given a more familiar name (Mike), and no mention was made of an alien planet.

Experiment 2 tested the role of perceptual similarity versus category information when considering familiar properties that are embedded in children's naïve theories of biology and human action—properties such as having a heart, having wires, breathing, or being sold in a

store (see Table 2). Our rationale for examining familiar properties was as follows. SKF argue that their data “[support] the idea that similarity-based induction is a default early in development, [and challenge] the idea of spontaneous category-based induction.” On a strong reading of this position, children should default to similarity rather than category information even for *familiar* properties, if tested with novel categories that decouple appearance and category information. In children’s prior experience, they have learned facts about particular animals and artifacts (that Fido has a heart, whereas their toaster has wires), but such generalizations could be represented as either similarity-based (furry things breathe, metallic things have wires) or as category-based (animals breathe, artifacts have wires). As SKF note: “... under more regular conditions, appearance information and category information are highly correlated, so that it is difficult to distinguish between similarity-based and category-based induction. Although children may have learned that typical animals (e.g., dogs) have hearts, and that typical machines (e.g., toasters) have wires inside, they have never encountered ziblets and flurps previously, and rarely (if ever) would they have encountered instances in which ontological categories’ membership is in direct conflict with perceptual features. Thus the current experiment provides a novel test of the role of categories vs. appearances in how children’s knowledge is organized.

Because we altered the perceptual similarity structure of the items (by adding identifying features: eyes vs. bolts), it was important to ensure that perceptual similarity alone could not yield the appropriate inferences. We therefore included a control condition that was identical to the experimental condition, except that the inference task came first, followed by the categorization task. If perceptual similarity alone is driving children’s inferences, then they should perform identically on the induction task, whether it comes before or after the categorization training. However, if categorization also contributes to children’s inferences, then they should make more category-based inferences in the experimental condition than in the control condition.

Method

Participants—Participants were 32 children (M age = 5.02, SD =0.47; 16 boys, 16 girls) and 32 adults (M age = 19.16, SD =1.05; 16 men, 16 women). Children were predominantly middle class and white. Adults were undergraduate students in an Introductory Psychology class. One additional child was dropped from the final data due to equipment malfunction. An additional 12 adults (6 men, 6 women; mean age 20.37) participated in pretesting of the properties. All participants were from a midsize city in the Midwestern United States.

Materials—Materials included the same items as in Experiment 1, with two changes. First, the 2 creatures presented during the category training phase of the experiment included one ziblet and one flurp (rather than 2 ziblets, as in Experiment 1). Second, for each ziblet, a circle was added to each antenna; for each flurp, an “X” was added to each antenna (see Figure 1b). Thus, eyes were always added to the creatures with more fingers than buttons, and bolts were always added to the items with more buttons than figures. The creatures and item sets were otherwise identical. We also included the drawing of Mike (Fritz from Experiment 1), as well as drawings of two contexts (tree, factory). The test properties differed from those in Experiment 1 (see Table 2). Half the test properties were appropriate for animals, and thus were assigned exclusively to target ziblets; half the test properties were appropriate for artifacts, and thus were assigned exclusively to target flurps.

In order to confirm that the test properties reflect ontologically meaningful knowledge, 12 adults participated in a pretest in which ziblets and flurps were first described (“Ziblets have eyes and eat grapes and live in trees”; “Flurps have bolts and use electricity and come from factories”), and then they were asked to rate how likely each property was to be true of

ziblets or flurps, on a scale of 1 (“ziblet only”) to 7 (“flurp only”). For example, they were asked to judge whether the property “has wires inside” was more likely to be true of ziblets or flurps. No pictures were provided, so that these ratings reflected pre-existing expectations based on property content alone. As predicted, the overall mean score for the animate properties ($M=2.08$) was significantly lower than the overall mean score for the artifact properties ($M=5.85$), $t_{\text{paired}}(11) = 5.40$, $p < .001$, thus confirming that, for adults, the properties are linked a priori to animals vs. artifacts.

Procedure—The procedure in the experimental condition differed from Experiment 1 in the following respects:

- the information provided during category training (see Table 1 and Appendix);
- the feedback during category learning (e.g., “Yes, it’s a ziblet. It has eyes and it goes in the tree” or “Yes, it’s a flurp. It has bolts and it goes in the factory”);
- the response options during category learning and the categorization tasks (both involved pointing to a habitat; e.g., “If it’s a ziblet, like this one, it goes there, in the tree. If it’s a flurp, like this one, it goes there, in the factory”);
- the addition of “eyes” and “bolts” to the ziblet and flurp stimuli (as described above; see also Figure 1b);
- the properties used during the induction task (as described in the Materials section; see also Table 2);
- no mention of pets, a pet store, or a pet store owner.

Additionally, Experiment 2 included a control condition in which participants received the tasks in a different order from the experimental condition (see Table 3). Specifically, those in the control condition received the induction task first, followed by category training, category learning, and initial categorization, in that order. Participants were randomly assigned to either the experimental condition or the control condition.

Results

First we tested to ensure that participants learned and remembered the relevant categories of “ziblets” and “flurps”. All participants passed the categorization criterion, and were highly accurate in the initial categorization task, with M s (out of 8) ranging from 7.94 (children) to 8.00 (adults), all greater than chance, p s $< .001$. Performance was also highly accurate in the final categorization task (experimental condition only), M s of 7.81 (children) and 8.00 (adults), both greater than chance, p s $< .001$.

The primary analyses focused on responses in the induction task. Each participant received a score indicating the number of trials (out of 12) on which they selected the categorical response (see Table 4). These scores were entered into a 2 (age group: children, adults) \times 2 (condition: experimental, control) univariate ANOVA, with both age group and condition as between-subjects factors. As predicted, there was a significant effect of condition, $F(1,60) = 76.27$, $p < .001$, $\eta_p^2 = .56$, indicating more category-based responses in the experimental condition ($M=10.66$) than the control condition ($M=4.66$). There was also a significant condition \times age group interaction, $F(1,60) = 9.57$, $p = .003$, $\eta_p^2 = .14$. Follow-up tests indicate that performance was higher in the experimental than control condition at both age groups, p s $< .001$. However, adults had significantly higher scores than children in the experimental condition (M s = 11.94 and 9.37, respectively), $p = .011$. Both children and adults selected the category-based response significantly above chance (of 6.0) in the experimental condition, p s $< .001$. In the control condition, children were at chance ($M=5.50$) and adults were below chance ($M=3.81$, $p = .024$).

Discussion

Experiment 2 demonstrates that when young children are asked to make inferences regarding novel items, they rely on category information more than appearances, *if*: (a) the categories are conceptually distinct (drawn from ontologically disparate domains), (b) the categories are distinguishable by readily accessible (albeit subtle) features that are biologically or mechanically plausible, and (c) the properties being inferred are familiar (e.g., can have babies) and known to be linked to typical exemplars of the relevant domains (e.g., dogs). This experiment thus demonstrates that children do not attend strictly to perceptual similarity when making inferences about novel items, thereby undermining a strict similarity-based position (“looks are everything”; Sloutsky et al., 2007).

One notable aspect of this experiment is that it examined children’s inferences regarding familiar properties (e.g., having a heart, having wires). Although the inferences were indisputably novel, in that children had never encountered ziblets or flurps before, we cannot know if children’s inferences were inductive (“This [target] ziblet can have babies; therefore, this [test] ziblet can have babies”) or deductive (“All animals can have babies; this [test] ziblet is an animal; therefore this [test] ziblet can have babies”). The task is thus importantly different from one in which children extend novel properties that were introduced for the first time in the experimental context.

Nonetheless, even if children were in fact making deductive inferences from the broader categories of “animal” and “thing”, the findings are important, as they provide new evidence regarding the role of categories and appearances in how children’s knowledge is organized. Although children have learned about prototypical animals and machines, an open question is how powerful a role these categories play when placed in conflict with appearances. Several prior studies have attempted to tease apart perceptual similarity and category information in children’s understanding of the biological domain (e.g., Carey, 1985; R. Gelman, 1990; S. Gelman & Nyhof, reported in Gelman, 2003; Jipson & Gelman, 2007; Massey & Gelman, 1988). Such work has examined preschool children’s attributions regarding items that look animate but are not (such as a doll, statue, or robotic pet), or items that look inanimate but are not (such as a starfish, echidna, or lettuce slug). However, in such studies there were substantial perceptual cues indicating the ontological kind of the target item (e.g., the metallic texture of a robotic pet; the self-initiated motion of a lettuce slug). The current experiment extends beyond prior work by examining a wholly novel set of items, for which membership in an ontological domain is strongly pitted against outward appearances, to determine which children judge to be more predictive.

A further difference between Experiments 1 and 2 is that the latter included unique perceptual features that distinguish ziblets and flurps (the markings on the antennae). Thus, one potential concern is that these added features may have changed the perceptual similarity structure of the induction triads, such that within-category similarity was greater than between-category similarity. The control condition was included in order to address this issue. In order to make certain that participants could not succeed on the induction task based on appearances alone, we reversed the order of the tasks in the control condition: the inference task came *first*, before children learned anything about the categories. We discovered, in this case, that the eyes and bolts were insufficient to cue either children or adults to draw the relevant inferences. Thus, it is not the perceptual similarity structure of the items that guides children’s performance, but rather the relevant categorical distinction.

Given that children clearly use category membership, not perceptual similarity, to guide their inferences regarding familiar biological and non-biological properties, the next step will be to see if this result extends to novel properties. This was the purpose of Experiment 3.

Experiment 3: Ontologically Distinct Categories, Novel Test Properties

Experiment 3 was identical to Experiment 2 in every way except for the test properties used in the induction task. The test properties were novel rather than familiar, in order to test whether children can use category information to guide novel inductive inferences. We were unable to use the novel test properties employed by SKF, as they were appropriate for animals only (e.g., “has a big brain”, “has round muscles”), and so could not apply to the flurps in Experiment 3. We therefore selected test properties that would be neutral regarding ontological type (e.g., unfamiliar functions, parts, behaviors), and could apply to either animate or inanimate items.

Method

Participants—Participants were 32 children (M age = 5.29, $SD=0.55$; 15 boys, 17 girls) and 32 adults (M age = 19.01, $SD=0.77$; 8 men, 24 women). Two additional children were tested but dropped, due to experimenter error. One additional adult was tested but not included because of language difficulties (non-native speaker who did not understand the task). None of the participants in the main task participated in Experiments 1 or 2. The 12 adults who participated in the property pretest in Experiment 2 also participated in a pretest of the properties in Experiment 3. Children were predominantly middle-class and white. Adults were undergraduate students in an Introductory Psychology class. All participants were from a midsize city in the Midwestern United States.

Materials—Materials were identical to those of Experiment 2, except for the test properties (see Table 2). Each test property was designed to be unfamiliar, and most included a novel word (e.g., “uses danner”). As in Experiment 2, half the test properties were assigned exclusively to target ziblets; half the test properties were assigned exclusively to target flurps. In order to confirm that the properties were not linked a priori to either ziblets or flurps, the same 12 adults who pretested the materials in Experiment 2 also rated how likely each property from Experiment 3 was to be true of ziblets or flurps, on a scale of 1 (“ziblet only”) to 7 (“flurp only”), after ziblets and flurps were first described (“Ziblets have eyes and eat grapes and live in trees”; “Flurps have bolts and use electricity and come from factories”). For example, they were asked to judge whether the property, “has blickets inside of it,” was more likely to be true of ziblets or flurps. No pictures were provided, so that these ratings reflected pre-existing expectations based on property content alone. As predicted, the overall mean score for the properties assigned to ziblets ($M=4.51$) did not significantly differ from the overall mean score for the properties assigned to flurps ($M=4.22$), paired- $t(11) = 1.53$, $p > .15$. (The small, non-significant difference obtained was in fact in the opposite direction to how the properties were assigned.) Thus, the pretest validates the properties as having no a priori link to ziblets vs. flurps, on the basis of content alone.

Procedure—The procedure was identical to that of Experiment 2 except for the induction properties, as described above.

Results

First we tested to ensure that participants learned and remembered the relevant categories of “ziblets” and “flurps”. All participants passed the categorization criterion, and were highly accurate in the initial categorization task, with M s (out of 8) of 7.91 (child), and 8.00 (adult), both greater than chance, p s $< .001$. Performance was also highly accurate in the final categorization task (experimental condition only), with M s of 7.06 (children) and 8.00 (adults), both greater than chance, p s $< .001$.

The primary analyses focused on responses in the induction task. Each participant received a score indicating the number of trials (out of 12) on which they selected the categorical response (see Table 4). These scores were entered into a 2 (age group: children, adults) x 2 (condition: experimental, control) univariate ANOVA, with both age group and condition as between-subjects factors. As predicted, there was a significant effect of condition, $F(1,60) = 36.62, p < .001, \eta_p^2 = .38$, indicating more category-based responses in the experimental condition ($M=8.47$) than the control condition ($M=3.75$). There was also a non-significant trend toward a condition x age group interaction, $F(1,60) = 3.55, p = .064, \eta_p^2 = .06$. Follow-up tests indicate that performance was higher in the experimental than control condition at both age groups, $p < .005$. However, adults had significantly lower scores than children in the control condition (M s = 2.63 and 4.88, respectively), $p < .05$. Both children and adults selected the category-based response significantly above chance (of 6.0) in the experimental condition, $p < .05$. In the control condition, children were non-significantly below chance ($p = .07$) and adults were significantly below chance ($p < .001$).

Discussion

Experiment 3 finds that when presented with novel categories from distinct ontological domains, with subtle perceptual cues indicating category membership (eyes vs. bolts), children and adults alike reliably make inductive inferences about novel properties on the basis of category membership. These findings extend beyond those of Experiment 2, in that the test properties were wholly novel. Thus, the properties were unfamiliar to participants and not linked to their prior knowledge about animals or artifacts. As in Experiment 2, these data cannot be due to the perceptual features of the stimuli alone, because participants did not select the category-based choices in a control condition in which the induction task was presented first, before they learned the conceptual basis of the categories.

Experiment 4: Basic-Level Natural Kinds, Novel Test Properties

Experiments 2 and 3 demonstrate that, for categories that contrast at an ontological level (ziblets are a type of animal; flurps are a type of artifact), children make category-based inferences regarding both familiar and novel properties. Ontological distinctions are cognitively special and arguably among the first concepts that children acquire (Carey, 2011; Keil, 1979). It is thus an open question as to whether categories that contrast at the basic level (e.g., two kinds of animals) would likewise elicit category-based inferences. Especially given that questions about the role of labeling in categorization often concern categories that contrast at the basic-level, this is an important issue to examine. Thus, Experiment 4 was designed to extend our research question to basic-level animal categories. Another important aspect of Experiment 4 is that we introduced wholly novel distinguishing features, for which children have no prior expectations or associations, in contrast to the eyes and bolts used in Experiments 2 and 3. Thus, to the extent that children continue to make use of categorical information, it cannot be due to prior associations with the features provided.

Method

Participants—Participants were 32 children (M age = 5.06, $SD=0.66$; 15 boys, 17 girls) and 32 adults (M age = 18.78, $SD=1.06$; 15 men, 17 women). An additional 9 children were tested but not included (7 didn't pass the 75% accuracy criterion on the initial categorization task; one was autistic; one experienced computer error). An additional 12 adults (4 men, 8 women; mean age 20.33) participated in a pretest of the properties in Experiment 4. Children were predominantly middle-class and white. Adults were undergraduate students in an Introductory Psychology class. All participants were from a midsize city in the Midwestern United States.

Materials—Materials were identical to those of Experiment 3, with three differences. First, the descriptions provided for ziblets vs. flurps contrasted at a basic level rather than an ontological level. Specifically, participants heard: “Ziblets have a round mouth and eat grapes and live in the jungle” and “Flurps have sharp teeth and eat nuts and live in the desert.” Second, although the distinguishing features were visually identical to those used in Experiments 2 and 3 (as shown in Figure 1b), they referred to two different kinds of mouths (round mouths, in the case of ziblets, and sharp teeth, in the case of flurps). And third, the pictorial context for flurps was a desert rather than a factory. The tree picture from Experiments 2–4 was used again as the jungle context.

In order to confirm that the properties were novel and thus not linked a priori to either ziblets or flurps, 12 adults participated in a pretest of the materials by rating how likely each property from Experiment 4 was to be true of ziblets (“Ziblets have a round mouth and eat grapes and live in the jungle”) or flurps (“Flurps have sharp teeth and eat nuts and live in the desert”), on a scale of 1 (“ziblet only”) to 7 (“flurp only”). No pictures were provided, so that these ratings reflected pre-existing expectations based on property content alone. As predicted, the overall mean score for the properties assigned to ziblets ($M=3.74$) did not significantly differ from the overall mean score for the properties assigned to flurps ($M=3.85$), $p > .56$. Thus, the pretest validates the properties as having no a priori link to ziblets vs. flurps, in the absence of pictures.

Procedure—The procedure was identical to that of Experiment 3 except for the information provided about how ziblets and flurps differ, as described above.

Results

First we tested to ensure that participants learned the relevant categories of ziblets and flurps. Participants who passed the categorization criterion were highly accurate in the initial categorization task, with no significant effects of age or conditions; M s (out of 8) = 7.84 (child), and 7.97 (adult), both greater than chance, $ps < .001$. Performance was also highly accurate in the final categorization task (experimental condition only), M s of 7.81 (children) and 8.00 (adults), both greater than chance, $ps < .001$.

The primary analyses focused on responses in the induction task. Each participant received a score indicating the number of trials (out of 12) on which they selected the categorical response (see Table 4). These scores were entered into a 2 (age group: children, adults) \times 2 (condition: experimental, control) univariate ANOVA, with both age group and condition as between-subjects factors. Importantly, there was a significant effect of condition, $F(1,60) = 27.54$, $p < .001$, $\eta_p^2 = .31$, indicating more category-based responses in the experimental condition ($M=8.47$) than the control condition ($M=4.12$). There was also a non-significant trend toward a condition \times age group interaction, $F(1,60) = 3.15$, $p = .081$, $\eta_p^2 = .05$. We therefore wished to determine whether the condition effect was upheld within each age group considered separately. Follow-up tests indicate that performance was higher in the experimental than control condition at both age groups, $ps < .02$. This result indicates that, for young children as well as adults, category-based inferences were significantly higher in the experimental condition (in which the conceptual basis of the category was provided) than in the control condition (in which the conceptual basis of the category was not provided). Furthermore, there were no significant differences between children or adults in either condition. Adults selected the category-based response significantly above chance (of 6.0) in the experimental condition, $M=9.06$, $p = .005$, and children showed a non-significant trend to respond above-chance, $M=7.88$, $p = .075$. In the control condition, children were at chance ($M=5.00$) and adults were significantly below chance ($M=3.25$, $p < .001$).

Discussion

In Experiment 4, both children and adults used basic-level category information to guide their inductive inferences. Specifically, they were more likely to generalize novel properties to an animal from the same basic-level kind (from one ziblet to another ziblet, or from one flurp to another flurp) in the experimental condition than in the control condition. In the experimental condition, they had learned the conceptual basis of the categories, and made use of them in their inductive inferences. However, in the control condition, despite the fact that the materials were perceptually identical to those in the experimental condition, there was no conceptual basis to the categories provided, and participants did not make use of them in their inductive inferences. These findings demonstrate a sensitivity to and use of category information even when the categorical distinction is at the basic level (as opposed to the ontological level, as tested in Experiments 2 and 3), and even when the distinguishing features are wholly novel (type of mouth and teeth; as opposed to familiar distinguishing features, as tested in Experiments 2 and 3).

Experiment 5: Non-Kind Categories with Additional Perceptual Features

Experiments 2, 3, and 4 demonstrate that children make category-based inferences for novel categories that contrast at either the ontological or the basic level, given sufficient conceptual information to indicate that the categories differ in fundamental respects (habitat and diet). This contrasts with performance in Experiment 1 and in SKF's original work, where children failed to use category information with categories of pets that were not as conceptually distinct. However, one important difference between Experiment 1 and Experiments 2–4 is that only the latter included perceptual cues that differentiate items from the two categories: ziblets and flurps actually differed in the markings on the antennae in Experiment 2–4, but not Experiment 1 (see Figures 1a vs. 1b). It thus could be argued that the perceptual cues per se were responsible for children's improved performance. In other words, perhaps children's inferences were driven by perceptual similarity, not category membership per se.

The major argument against this interpretation is that children (and adults) performed poorly in the control conditions of Experiments 2–4, in which the same perceptual cues were present, but no category information was provided. Based on the control conditions, we can conclude that the perceptual information provided by the test stimuli were insufficient to yield a pattern of category-based inferences. That is, simply seeing the distinguishing features was not enough to cue participants to make use of them on the induction task, as those in the control condition viewed the features but did not make category-based inductive inferences.

The goal of Experiment 5 is to provide further evidence to determine whether the category-based performance in Experiments 2–4 could be due to the perceptual cues from the addition of the subtle markings on the antennae. One possibility not tested in the prior studies is that children may make use of the additional features when they are associated with the categorical distinction of ziblets vs. flurps. When such features are not linked to the ziblet-flurp distinction, children may have simply ignored the additional features. Thus, in Experiment 5, all stimuli included the added perceptual cues on the antennae (as in Figure 1b), and additionally we provided training on the ziblet-flurp category distinctions used in Experiment 1 (and in the original SKF work). In other words, children received training and labels about ziblets and flurps, while simultaneously having access to the additional differentiating features (Os and Xs on the antennae). As in Experiment 1, the SKF script was used (i.e., the categorization rule concerned the finger:button ratio; no information was provided regarding diet or habitat; test questions concerned SKF's novel test properties). If children's success in Experiments 2–4 was due to the availability of a visible perceptual cue

differentiating ziblets and flurps, then children should make category-based inductive inferences in Experiment 5. However, if children's success in Experiments 2–4 was due to conceptual information, then the results of Experiment 5 should replicate those of Experiment 1. That is, children should fail to make category-based inductive inferences in Experiment 5, just as they did in Experiment 1.

Method

Participants—Participants were 32 children (M age = 5.00, SD =0.63; 17 boys, 15 girls) and 32 adults (M age = 19.06, SD =1.01; 13 men, 19 women). Seven additional children were tested but dropped (4 did not meet the 75% accuracy criterion for initial categorization; one was not a fluent speaker of English; two were below the target age range). Children were predominantly middle-class and white. Adults were undergraduate students in an Introductory Psychology class. All participants were from a midsize city in the Midwestern United States.

Materials—The pictures of ziblets and flurps were identical to those used in Experiments 2–4. All other materials were identical to those of Experiment 1.

Procedure—Participants were randomly assigned to either an experimental condition or a control condition. The procedure of the experimental condition was identical to that of Experiment 1. The procedure of the control condition was identical to that of the experimental condition, except that the tasks were in modified order and there was no final categorization task (see Table 3).

Results

First we tested to ensure that participants learned the relevant categories of ziblets and flurps. Participants who passed the initial categorization criterion were highly accurate in the initial categorization task, although there was a main effect of age group, $F(1,60) = 8.12$, $p = .006$, $\eta_p^2 = .12$, indicating that adults performed better than children; M s (out of 8) = 7.59 (child), 7.97 (adult), both greater than chance, p s < .001. Performance was also highly accurate in the final categorization task (experimental condition only), though again adults performed better than children (M s of 6.88 children, 8.00 adults), $p < .02$, both greater than chance, p s < .001.

The primary analyses focused on responses in the induction task. Each participant received a score indicating the number of trials (out of 12) on which they selected the categorical response (see Table 4). These scores were entered into a 2 (age group: children, adults) x 2 (condition: experimental, control) univariate ANOVA, with both age group and condition as between-subjects factors. In contrast to Experiments 2–4, we obtained no significant effect of condition, $p > .13$. There was a marginal effect of age group, $F(1,60) = 3.66$, $p = .061$, $\eta_p^2 = .06$. Most importantly, we obtained a significant condition x age group interaction, $F(1,60) = 14.03$, $p < .001$, $\eta_p^2 = .19$. Follow-up tests indicate that adults performed better in the experimental than control condition, $p < .001$. In contrast, children showed no significant difference between the control and experimental conditions, $p = .12$, with the mean actually higher in the control condition. There were no significant differences between children or adults in the control condition, $p = .20$, but adults scored significantly higher than children in the experimental condition, $p < .01$. Children selected the category-based response significantly below chance (of 6.0) in the experimental condition, $M=3.25$, $p = .001$, and adults did not differ from chance, $M=7.69$, $p = .17$. In the control condition, children were at chance ($M=5.00$) and adults were significantly below chance ($M=3.56$, $p < .001$).

Discussion

Experiment 5 demonstrates that markings on the antennae did not alter the perceptual structure of items sufficiently to foster category-based responses. Even though the Xs and Os perfectly correlated with membership in the ziblet vs. flurp category (i.e., the creatures with more fingers than buttons always had Os; the creatures with more buttons than fingers always had Xs), and participants viewed these cues on 16 trials during category learning and initial categorization, this was insufficient to prime use of these cues during induction. This finding suggests that something other than perceptual features of the stimuli per se was responsible for their category-based induction in Experiments 2–4.

An interesting unanticipated result was that, for children, performance was actually slightly (though non-significantly) worse in the experimental condition, compared to the control condition. This pattern suggests that the added load of learning a complex categorization may actually present an obstacle to making use of the categories provided.

Experiment 6: Non-Kind Categories with Additional Perceptual and Contextual Features

Experiment 5 demonstrated that perceptual features of the stimuli in Experiments 2–4 were insufficient to prompt category-based responses. However, there was an additional perceptual component provided in those studies that may have contributed to children's improved performance, namely, the linking of the pictures to distinct images during initial categorization. Recall that the habitat pictures (Experiments 2 and 3: tree vs. factory; Experiment 4: tree [representing jungle] vs. desert) were provided to help reinforce the conceptual basis of the categorical distinction. Although these pictures were not provided during the induction test, and thus could not have altered the perceptual cues during testing, nonetheless, it is possible that perceptual associations learned during initial categorization may have been sufficient to cue use of the category.

Experiment 6 was designed to test this possibility. In this experiment, the initial categorization task was structurally identical to that of Experiments 2–4, in that the two contrasting categories (ziblets vs. flurps) were associated with two distinct perceptual images. Specifically, during initial categorization, participants were instructed to point to either a purple rectangle (to indicate a ziblet) or a brown rectangle (to indicate a flurp). From a perceptual analysis, this is equivalent to the sorting task provided in Experiments 2–4, in that ziblets and flurps are each correlated with an overt perceptual cue. From a conceptual standpoint, however, the linking is arbitrary and thus should not help with induction. Because adults performed well across the board in the prior experiments, we included only children in this experiment.

Method

Participants—Participants were 32 children (M age = 4.96, $SD=0.67$; 16 boys, 16 girls). Two additional children were tested but dropped (one for not being a fluent speaker of English; the other for failing to meet the 75% accuracy criterion during initial categorization). Children were predominantly middle-class and white. All participants were from a midsize city in the Midwestern United States.

Materials—One purple rectangle and one brown rectangle (identical in size to the habitat pictures in Experiments 2–4; both with marbled texture) were used during the sorting task. All other materials were identical to those of Experiment 5, including the distinguishing features provided on the antennae (Os vs. Xs).

Procedure—The procedure was identical to that of Experiment 5, except that during initial and final categorization, participants were given the option of pointing to a purple versus brown rectangle to indicate a ziblet or a flurp.

Results

First we tested to ensure that children learned the relevant categories of ziblets and flurps. Those who met the categorization criterion were highly accurate in the initial categorization task; M (out of 8) = 7.66, greater than chance, $p < .001$. Performance was also highly accurate in the final categorization task (experimental condition only), $M=6.94$, greater than chance, $p < .001$.

The primary analyses focused on responses in the induction task. Each participant received a score indicating the number of trials (out of 12) on which they selected the categorical response (see Table 4). We conducted a simple t -test comparing performance in the experimental and control conditions. In contrast to Experiments 2–4, and consistent with Experiments 1 and 5, children showed no significant difference between the control and experimental conditions, $p > .8$. Children selected the category-based response significantly below chance (of 6.0) in both the experimental condition, $M=4.00$, $p < .01$, and the control condition, $M=4.13$, $p = .001$.

Discussion

Experiment 6 was designed to be structurally identical to Experiments 2–4, in which perceptually and spatially distinct sorting locations were provided during the initial categorization phase of the experiment. The purpose of this procedural modification was to test whether including perceptual correlates in the context during categorization would affect performance later in an induction task. Despite these additional cues, children in Experiment 6 showed no evidence of using the category in their inductive inferences. Performance was significantly below chance, and there was no significant difference between the experimental condition (in which the categories were introduced before the induction task) and the control condition (in which the categories were introduced after the induction task). These findings imply that, in Experiments 2–4, children’s use of the category during induction was not based on contextual cues during categorization, and further that conceptual information (beyond the perceptual features of the task) influenced children’s use of the category during induction.

Experiment 7: Non-Kind Categories with Simpler Distinguishing Features

Experiments 5 and 6 demonstrated that perceptual features of the stimuli in Experiments 2–4 (Os and Xs on the antennae) were insufficient to prompt category-based responses. However, in those experiments, unlike Experiments 2–4, children were not trained to attend to the added perceptual features. It may be that training is required for children to notice and use them. We therefore designed Experiment 7 to provide training on the added perceptual distinction, but otherwise maintaining the original SKF categories and procedure from Experiment 5. If the conceptual account is correct, then children should still make perceptually-based inductions, because the categories and procedure provide a conceptually weak basis for induction (subordinate-level distinction, use of a non-biological distinguishing feature, lack of rich associated features, etc.), as discussed in the Introduction. However, if the addition of the antennae markings and training to notice this perceptual feature increase the perceived similarity within the flurp/ziblet categories, then children should make more “category”-based inductions, similar to children in Experiments 2–4, although these responses would actually be driven by perceptual similarity. Because adults

performed well across the board in the prior experiments, we included only children in this experiment.

Method

Participants—Participants were 16 children (M age = 4.67, $SD=0.35$; 8 boys, 8 girls). Two additional children were tested but dropped for failing to meet the 75% accuracy criterion during initial categorization. Children were predominantly middle-class and white. All participants were from a midsize city in the Midwestern United States.

Materials—All materials were identical to those of Experiment 5, including the distinguishing features provided on the antennae (Os vs. Xs).

Procedure—The procedure was identical to that of Experiment 5, except that children were instructed that the features distinguishing ziblets and flurps were the marks on the antennae (circles vs. Xs; see Appendix for wording). In addition, a sample ziblet was shown before final categorization, as a reminder (although no mention was made of the distinguishing features).

Results

First we tested to ensure that children learned the relevant categories of ziblets and flurps. Those who met the categorization criterion were highly accurate in the initial categorization task; M (out of 8) = 7.69, greater than chance, $p < .001$. Performance was also highly accurate in the final categorization task (experimental condition only), $M=6.13$, greater than chance, $p = .011$.

The primary analyses focused on responses in the induction task. Each participant received a score indicating the number of trials (out of 12) on which they selected the categorical response (see Table 4). Children selected the category-based response no greater than chance (of 6.0), $M=5.56$, $p > .6$. We also conducted a simple t -test comparing performance in this experimental condition with performance in the control condition of Experiment 5 (which used the same items and procedure for the induction task).^{iv} In contrast to Experiments 2–4, and consistent with Experiments 1, 5, and 6, children showed no significant difference between the control and experimental conditions, $p > .6$.

Discussion

Experiment 7 was designed to focus children's attention on the distinguishing features used in Experiments 2–4, to test whether these perceptual features could account for the greater use of categories in Experiments 2–4. Despite these additional cues, children in Experiment 7 showed little evidence of using the category in their inductive inferences. Performance was no different from chance, and there was no significant difference between the experimental condition (in which the categories were introduced before the induction task) and the control conditions of Experiments 5 or 6 (in which the categories were introduced after the induction task). These findings imply that, in Experiments 2–4, children's use of the category during induction was not based on perceptual cues present during categorization and test.

^{iv}The same non-significant results were obtained when using the Control condition from Experiment 6 as comparison.

General Discussion

In the present series of experiments, 4- and 5-year-old children and adults learned novel categories (ziblets and flurps) that were defined on the basis of a subtle rule. After sufficient training to learn the rule thoroughly, they were then asked to perform an inductive inference task in which category information and perceptual information were placed in conflict. The categories were not explicitly mentioned during testing (i.e., no labels were provided). Replicating previous research by Sloutsky et al. (2007), in Experiment 1 we found that children based their inferences on perceptual similarity, not category membership, when the novel categories were not natural kinds (Experiments 1, 5, 6, and 7). That is, when ziblets and flurps were two kinds of pets differing only in personality (nice and friendly versus mean and dangerous), children failed to use the categories as the basis for their inductive inferences. We also discovered that children differed from adults in this respect. In contrast to young children, adults consistently made use of the category at above-chance levels.

These findings taken in isolation could support either of two conclusions: (1) that there are stable, across-the-board developmental differences in the use of categorical versus perceptual information, with children focused on perceptual cues and adults focused on categorical cues, or (2) that children are flexible in their inductive inferences, and would make use of categorical information if it were more predictive and conceptually rich. In order to tease apart these two possibilities, we modified the items and procedure in order to examine performance for categories that are conceptually more distinct and more predictive than those used in Experiment 1 and prior research. These changes were of four types: (a) We presented categories that were conceptually more distinct: from different ontological categories [Experiments 2 and 3], or from different basic-level categories [Experiment 4], rather than different kinds of pets that may contrast at the subordinate level, (b) we emphasized the conceptual basis of the categories in the training phase and in the categorization tasks, and (c) we provided subtle but easily accessible cues to category membership. Furthermore, (d) we modified the test properties so that they assessed familiar properties (Experiment 2) or novel, non-obvious features (Experiments 3 and 4), rather than body parts that could be construed as linked to visible characteristics (e.g., round muscles).

We discovered that, in contrast to Experiment 1, when participants learned categories that have a deeper conceptual basis (Experiments 2–4), children like adults made use of category membership more than perceptual similarity as the basis of their inferences. This result holds for novel properties (e.g., “has blickets inside”) as well as familiar properties (e.g., “has wires inside”), and for novel distinguishing characteristics (e.g., different kinds of mouths) as well as familiar distinguishing characteristics (e.g., eyes vs. bolts). Differences in performance between Experiment 1 and Experiments 2–4 cannot be due to differences in how well participants learned the categories of ziblets and flurps, because in all studies with both age groups, performance on the categorization tasks was excellent, both before the induction task and at the very end of the experimental session. Participants’ performance also cannot be due to perceptual features of the test items (including the added Os and Xs on the antennae), as the visual aspects of the stimuli were identical in Experiments 2–7, and children were trained to attend to these features in Experiment 7. Finally, participants’ performance cannot be due to the response option of pointing to visual images (e.g., tree vs. factory) during the categorization tasks, as Experiment 6 rules this out. Altogether, then, these findings demonstrate that children are capable of using categorical information as well as perceptual similarity to guide their inductions, thus arguing against the perceptual similarity account and in favor of the conceptual account.

Although the perceptual similarity account is unsupported by these experiments, an open question is whether children’s judgments might still be characterized in terms of similarity

more broadly construed. Adding conceptually meaningful properties (such as diet and habitat) modified the overall similarity among items, by altering the number of shared or distinctive features, even if it did not alter the *perceptual* similarity among items. It may be that a broader similarity model could account for the present data (see Hayes, Heit, & Swendsen, 2010, for examples). However, if features are differentially weighted according to their conceptual centrality (e.g., internal features weighted more heavily than external features; ontological information weighted more heavily than other features), this may indicate that similarity per se is insufficient to account for people's judgments, and that additional factors (e.g., causal beliefs, essence placeholders, or a special status for natural kind labels) are required to account for judgments of feature centrality.

The present results seem to suggest that children differentially weighted different types of features (compare, for example, Experiments 4 and 6, in which the sheer number of features is comparable across the studies, including visual images associated with each category, but the results are strikingly different). Such a pattern is consistent with prior research finding that weighting of features is driven by domain-specific beliefs, causal reasoning, intuitive theories, and other considerations, for both children and adults (e.g., Ahn & Luhmann, 2005; Hayes & Thompson, 2007; Keil, 1989; Medin & Shoben, 1988; Murphy & Medin, 1985; Newman, Hermann, Wynn, & Keil, 2008; Rehder, 2009; Sloman, Love, & Ahn, 1998).

Another important finding from these experiments is that labeled categories are not all equivalent; they differ substantially from one another in their inductive potential. This finding runs counter to the perceptual similarity account, which posits domain-general effects of visual and auditory similarity, and thus does not predict any differences due to category content. In contrast, variability among categories fits with the conceptual account, which argues that children make use of categories (and labels) to the extent that they are meaningfully predictive of important features (see also Opfer & Bulloch, 2007). The present results are also consistent with prior research showing that different kinds of labels support different kinds of inferences. For example, Graham, Booth, and Waxman (2012) found that common noun labels support category-based inferences but adjectives support appearance-based inferences. Children are also more likely to make inferences based on: basic-level labels versus superordinate-level labels (Gelman & O'Reilly, 1988), labels that map onto perceptually coherent categories versus labels that do not (Davidson & Gelman, 1990), and labels that map onto causally rich categories versus labels that do not (Booth, 2012). The idea that labeled categories are not all equivalent also fits with work with adults, finding stable and sizeable distinctions between different kinds of categories (Ahn, Taylor, Kato, Marsh, & Bloom, in press; Medin, Lynch, & Solomon, 2000; Prasada, Hennefield, & Otap, 2012; Smith, Patalano, & Jonides, 1998; Yamauchi & Markman, 1998).

From the current set of experiments, we cannot isolate what precisely helped children make more category-based inferences, as the modifications in Experiments 2–4 entailed several components (outlined in Table 3), including: a more conceptually based categorical distinction (ontological- or basic-level kinds), kind-relevant correlated features (diet and habitat), contrasting exemplars during training to highlight the categorical contrasts, cues to categorization that are biologically and functionally meaningful (subtly contrasting parts, embedded in the antennae), test properties that assessed category-relevant, non-obvious features (e.g., non-obvious internal parts, behaviors, and functions), and more child-friendly language. It may be that one, some, or all of these changes affected performance.

Nonetheless, our view is that the key features are likely those that serve to signal that categories are conceptually rich natural kinds. As noted in the introduction, at least four factors may contribute to making categories more conceptually distinct: the level at which categories differ from one another; the richness of shared features within a category;

distinguishing features that are inherent and functional parts; and emphasizing contrasts between different categories. We hypothesize that each of these factors provides meaningful information to children as they learn to discern which categories and labels are inference-promoting (e.g., marmosets, copper) and which are not (e.g., containers, pets). In future research it would be valuable to test which of these factors are sufficient to yield category-based inferences, and whether some are more important than others.

An important further aspect of these data concerns the changes taking place between early childhood and young adulthood. Most strikingly, adults made many more category-based inferences than children when considering the pet categories developed by SKF (Experiments 1 and 5). These data are consistent with prior work showing that, in certain contexts, young children are more likely than adults to use perceptual similarity to guide their inductive inferences, and conversely adults are more likely than children to use category membership to guide their inductive inferences (Badger & Shapiro, 2012; Deng & Sloutsky, 2012, 2013). Also notable, however, are that these developmental differences decreased or even disappeared in Experiments 2–4. Altogether, then, the data suggest that what primarily changes with age is the information that is used when reasoning about categories whose kind status is unclear. We found that for categories that lack the signature features of natural kinds (e.g., characteristic and distinctive diet, habitat, locomotion, vocalizations), and thus are ambiguous in their status, children place greater weight on perceptual information to guide their inductive inferences, whereas adults place greater weight on language (specifically, noun labels) as a cue. Both perceptual features and labels are generally good correlates of kind membership and thus are sensible cues (Diesendruck & Bloom, 2003; see Hayes, 2007, for similar arguments regarding the intercorrelations among similarity, causal structure, and categorical relations). However, prior work has shown that children become increasingly sensitive to cue validity in their inferences (Bulloch & Opfer, 2009). In the case of category-based induction, this developmental change may manifest as children showing a more conservative interpretation of labels than adults, whereby they need more evidence that a label refers to a richly structured kind before they are willing to rely on it to guide their inferences.

In closing, we note that the distinction between perceptual and conceptual accounts of how labels influence inductive inferences can be understood as part of a broader set of classic tensions concerning the nature of cognitive development. Waxman and Gelman (2009) note that early language and cognitive development are often discussed in terms of two metaphors, emphasizing the child as “data-analyst” (with a focus on children’s impressive capacity to attend to and organize statistical environmental cues) or the child as “theorist” (with a focus on children’s equally impressive capacity to engage in causal reasoning to construct larger knowledge structures). They suggest that the most sensible account is one that marries these two metaphors, acknowledging both the data-analytic capacities and the theory-building capacities of young children. The present findings support precisely such a view.

Acknowledgments

This research was supported by NICHD grant HD-36043 to Gelman. We thank the University of Michigan Language Lab for helpful discussion. We are grateful to the children and parents who participated in the research, and thank Dayana Kupisk, Lisa Chen, Megan Helfend, and Gladys Tan for their assistance. We would also like to thank the following preschools: Annie’s Children’s Center, Children’s Creative Center, Childtime, Discovery Center, Generations Together, Little Folks Corner, Tutor Time, and the University of Michigan’s Towsley Center.

References

- Ahn, W.; Luhmann, CC. Demystifying theory-based categorization. In: Gershkoff-Stowe, L.; Rakison, D., editors. *Building object categories in developmental time*. Mahwah, NJ: Erlbaum; 2005. p. 277-300.
- Ahn W, Taylor EG, Kato D, Marsh J, Bloom P. Causal essentialism in kinds. *Quarterly Journal of Experimental Psychology*. (in press).
- Akhtar, N.; Tomasello, M. The social nature of words and word learning. In: Golinkoff, R.; Hirsh-Pasek, K., editors. *Becoming a word learner: A debate on lexical acquisition*. New York: Oxford University Press; 2000.
- Badger JR, Shapiro LR. Evidence of a transition from perceptual to category induction in 3- to 9-year-old children. *Journal of Experimental Child Psychology*. 2012;1016/j.jecp.2012.03.004
- Baldwin DA, Markman EM, Melartin RL. Infants' ability to draw inferences about nonobvious object properties: Evidence from exploratory play. *Child Development*. 1993; 64(3):711–728.10.2307/1131213 [PubMed: 8339691]
- Booth, AE. Unpublished ms. Northwestern University; 2012. Causally rich categories support name-based inductive inference in preschoolers.
- Bulloch MJ, Opfer JE. What makes relational reasoning smart? Revisiting the perceptual-to-relational shift in the development of generalization. *Developmental Science*. 2009; 12(1):114–122.10.1111/j.1467-7687.2008.00738.x [PubMed: 19120419]
- Carey, S. *Conceptual change in childhood*. Cambridge, MA: Bradford Books, MIT Press; 1985.
- Carey, S. *The origin of concepts*. New York: Oxford University Press; 2009.
- Davidson NS, Gelman SA. Inductions from novel categories: The role of language and conceptual structure. *Cognitive Development*. 1990; 5(2):151–176.10.1016/0885-2014(90)90024-N
- Deng W, Sloutsky VM. Carrot eaters or moving heads: Inductive inference is better supported by salient features than by category labels. *Psychological Science*. 2012; 23(2):178–186. [PubMed: 22228644]
- Deng WW, Sloutsky VM. The role of linguistic labels in inductive generalization. *Journal of Experimental Child Psychology*. 2013; 114(3):432–55.10.1016/j.jecp.2012.10.011 [PubMed: 23270793]
- Dewar K, Xu F. Do early nouns refer to kinds or distinct shapes? Evidence from 10-month-old infants. *Psychological Science*. 2009; 20(2):252–257.10.1111/j.1467-9280.2009.02278.x [PubMed: 19175526]
- Diesendruck G, Bloom P. How specific is the shape bias? *Child Development*. 2003; 74(1):168–178.10.1111/1467-8624.00528 [PubMed: 12625443]
- Gelman R. First principles organize attention to and learning about relevant data: Number and the animate-inanimate distinction as examples. *Cognitive Science*. 1990; 14(1):79–106.10.1207/s15516709cog1401_5
- Gelman SA. The development of induction within natural kind and artifact categories. *Cognitive Psychology*. 1988; 20(1):65–95.10.1016/0010-0285(88)90025-4 [PubMed: 3338268]
- Gelman, SA. *The essential child: Origins of essentialism in everyday thought*. New York, NY: Oxford University Press; 2003.
- Gelman SA, Ebeling KS. Shape and representational status in children's early naming. *Cognition*. 1998; 66(2):B35–B47.10.1016/S0010-0277(98)00022-5 [PubMed: 9677767]
- Gelman SA, Markman EM. Categories and induction in young children. *Cognition*. 1986; 23(3):183–209.10.1016/0010-0277(86)90034-X [PubMed: 3791915]
- Gelman SA, Markman EM. Young children's inductions from natural kinds: The role of categories and appearances. *Child Development*. 1987; 58(6):1532–1541.10.2307/1130693 [PubMed: 3691200]
- Gelman SA, Medin DL. What's so essential about essentialism? A different perspective on the interaction of perception, language, and conceptual knowledge. *Cognitive Development*. 1993; 8(2):157–167.10.1016/0885-2014(93)90011-S
- Gelman SA, O'Reilly AW. Children's inductive inferences within superordinate categories: The role of language and category structure. *Child Development*. 1988; 59(4):876–887.10.2307/1130255 [PubMed: 3168626]

- Gelman SA, Waxman SR. Looking beyond looks: Comments on Sloutsky, Kloos, and Fisher (2007). *Psychological Science*. 2007; 18(6):554–555.10.1111/j.1467-9280.2007.01937.x [PubMed: 17576270]
- Graham SA, Kilbreath CS, Welder AN. Thirteen-month-olds rely on shared labels and shape similarity for inductive inferences. *Child Development*. 2004; 75(2):409–427.10.1111/j.1467-8624.2004.00683.x [PubMed: 15056196]
- Hayes, BK. The development of inductive reasoning. In: Feeney, A.; Heit, E., editors. *Inductive reasoning: Experimental, developmental, and computational approaches*. New York: Cambridge University Press; 2007. p. 25-54.
- Hayes BK, Heit E, Swendsen H. Inductive reasoning. *Wiley Interdisciplinary Reviews: Cognitive Science*. 2010; 1:278–292.
- Hayes BK, Thompson SP. Causal relations and feature similarity in children’s inductive reasoning. *Journal of Experimental Psychology: General*. 2007; 136(3):470–484.10.1037/0096-3445.136.3.470 [PubMed: 17696694]
- Jipson JL, Gelman SA. Robots and rodents: Children’s inferences about living and nonliving kinds. *Child Development*. 2007; 78(6):1675–1688.10.1111/j.1467-8624.2007.01095.x [PubMed: 17988314]
- Keil, FC. *Semantic and conceptual development: An ontological perspective*. Cambridge, MA: Harvard University Press; 1979.
- Keil, FC. *Concepts, kinds, and cognitive development*. Cambridge, MA: The MIT Press; 1989.
- Leslie SJ. Generics: Cognition and acquisition. *Philosophical Review*. 2008; 117:1–47.
- Lupyan G, Rakison DH, McClelland JL. Language is not just for talking: Redundant labels facilitate learning of novel categories. *Psychological Science*. 2007; 18(12):1077–1083.10.1111/j.1467-9280.2007.02028.x [PubMed: 18031415]
- Massey CM, Gelman R. Preschooler’s ability to decide whether a photographed unfamiliar object can move itself. *Developmental Psychology*. 1988; 24(3):307–317.10.1037/0012-1649.24.3.307
- Medin DL, Lynch EB, Solomon KO. Are there kinds of concepts? *Annual Review of Psychology*. 2000; 51:121–147.10.1146/annurev.psych.51.1.121
- Medin DL, Shoben EJ. Context and structure in conceptual combination. *Cognitive Psychology*. 1988; 20(2):158–190.10.1016/0010-0285(88)90018-7 [PubMed: 3365938]
- Mill, JS. *A system of logic, ratiocinative and inductive*. London: Longmans; 1843.
- Moely BE, Olson FA, Halwes TG, Flavell JH. Production deficiency in young children’s clustered recall. *Developmental Psychology*. 1969; 1(1):26–34.10.1037/h0026804
- Murphy, GL. *The big book of concepts*. Cambridge, MA: MIT Press; 2002.
- Murphy GL, Medin DL. The role of theories in conceptual coherence. *Psychological Review*. 1985; 92(3):289–316.10.1037/0033-295X.92.3.289 [PubMed: 4023146]
- Namy LL, Clepper LE. The differing roles of comparison and contrast in children’s categorization. *Journal of Experimental Child Psychology*. 2010; 107(3):291–305.10.1016/j.jecp.2010.05.013 [PubMed: 20609449]
- Newman GE, Herrmann P, Wynn K, Keil FC. Biases towards internal features in infants’ reasoning about objects. *Cognition*. 2008; 107(2):420–432.10.1016/j.cognition.2007.10.006 [PubMed: 18068151]
- Nisbett RE, Krantz DH, Jepson C, Kunda Z. The use of statistical heuristics in everyday inductive reasoning. *Psychological Review*. 1983; 90(4):339–363.10.1037/0033-295X.90.4.339
- Noles NS, Gelman SA. Effects of categorical labels on similarity judgments: A critical analysis of similarity-based approaches. *Developmental Psychology*. 2012a; 48(3):890–896.10.1037/a0026075 [PubMed: 22059447]
- Noles NS, Gelman SA. Disentangling similarity judgments from pragmatic judgments: Response to Sloutsky and Fisher (2012). *Developmental Psychology*. 2012b; 48(3):901–906.10.1037/a0027831 [PubMed: 22545856]
- Noles NS, Gelman SA. Preschool-age children and adults flexibly shift their preferences for auditory versus visual modalities but do not exhibit auditory dominance. *Journal Of Experimental Child Psychology*. 2012c; 112(3):338–350.10.1016/j.jecp.2011.12.002 [PubMed: 22513210]

- Opfer JE, Bulloch MJ. Causal relations drive young children's induction, naming, and categorization. *Cognition*. 2007; 105(1):206–217.10.1016/j.cognition.2006.08.006 [PubMed: 17045580]
- Osherson DN, Smith EE, Wilkie O, López A, Shafir E. Category -based induction. *Psychological Review*. 1990; 97(2):185–200.10.1037/0033-295X.97.2.185
- Prasada S, Hennefield L, Otap D. Conceptual and linguistic representations of kinds and classes. *Cognitive Science*. 2012; 36:1224–1250. [PubMed: 22671567]
- Rehder B. Casual-based property generalization. *Cognitive Science*. 2009; 33(3):301–344.10.1111/j.1551-6709.2009.01015.x [PubMed: 21585473]
- Rhodes M, Gelman SA. Categories influence predictions about individual consistency. *Child Development*. 2008; 79(5):1270–1287.10.1111/j.1467-8624.2008.01188.x [PubMed: 18826525]
- Rips LJ. Inductive judgments about natural categories. *Journal of Verbal Learning and Verbal Behavior*. 1975; 14(6):665–681.10.1016/S0022-5371(75)80055-7
- Rosch E. Basic objects in natural categories. *Cognitive Psychology*. 1976; 8(3):382–439.10.1016/0010-0285(76)90013-X
- Schwartz, SP., editor. *Naming, necessity, and natural kinds*. Ithaca, NY: Cornell University Press; 1977.
- Schwartz SP. Natural kind terms. *Cognition*. 1979; 7:301–315.
- Shiple, EF. Categories, hierarchies, and induction. In: Medin, DL., editor. *The psychology of learning and motivation*. San Diego, CA: Academic Press; 1993. p. 265-301.
- Sloman SA, Love BC, Ahn W. Feature centrality and conceptual coherence. *Cognitive Science*. 1998; 22(2):189–228.10.1016/S0364-0213(99)80039-1
- Sloutsky VM, Fisher AV. Induction and categorization in young children: A similarity-based model. *Journal of Experimental Psychology: General*. 2004; 133(2):166–188.10.1037/0096-3445.133.2.166 [PubMed: 15149249]
- Sloutsky VM, Fisher AV. Effects of categorical labels on similarity judgments: A critical evaluation of a critical analysis: Comment on Noles and Gelman (2012). *Developmental Psychology*. 2012; 48(3):897–900.10.1037/a0027531 [PubMed: 22545855]
- Sloutsky VM, Kloos H, Fisher AV. When looks are everything: Appearance similarity versus kind information in early induction. *Psychological Science*. 2007; 18(2):179–185.10.1111/j.1467-9280.2007.01869.x [PubMed: 17425540]
- Smith EE, Patalano AL, Jonides J. Alternative strategies of categorization. *Cognition*. 1998; 65(2–3):167–196.10.1016/S0010-0277(97)00043-7 [PubMed: 9557382]
- Taylor MG, Gelman SA. Children's gender- and age-based categorization in similarity and induction tasks. *Social Development*. 1993; 2(2):104–121.10.1111/j.1467-9507.1993.tb00006.x
- Waxman, SR. Everything had a name, and each name gave birth to a new thought: Links between early word learning and conceptual organization. In: Hall, D.; Waxman, SR., editors. *Weaving a lexicon*. Cambridge, MA: MIT Press; 2004. p. 295-335.
- Waxman SR, Gelman SA. Early word-learning entails reference, not merely associations. *Trends in Cognitive Sciences*. 2009; 13(6):258–263.10.1016/j.tics.2009.03.006 [PubMed: 19447670]
- Waxman SR, Lynch EB, Casey K, Baer L. Setters and samoyeds: The emergence of subordinate level categories as a basis for inductive inference in preschool-age children. *Developmental Psychology*. 1997; 33(6):1074–1090.10.1037/0012-1649.33.6.1074 [PubMed: 9383629]
- Yamauchi T, Markman AB. Category learning by inference and classification. *Journal of Memory and Language*. 1998; 39(1):124–148.10.1006/jmla.1998.2566

APPENDIX – Scripts for category training portion of each experiment

Experiments 1, 5, 6, and 7

This is Fritz. He lives on the planet Elbee with his family, and he would like to get a pet. There are different kinds of animals on the planet and they often look alike. Can you help him find a pet?

Here is the store on the planet Elbee. The problem with this store is that there are two different kinds of animals in the store that look alike, ziblets and flurps. Ziblets and flurps are very different: ziblets are nice and friendly animals and they are good pets. Flurps are wild and dangerous animals, and they do not make good pets. Even though they look a lot like ziblets, they are mean animals that nobody wants as a pet. We have to make sure Fritz finds a ziblet in the pet store and not one of the flurps. Can you help him find a ziblet?

(Experiments 1, 5, & 6 only)

Here are the body parts of animals in the pet store. They have a body, antennas, a tail, wings, buttons, and fingers on each wing. And they eat insects that get stuck on their body. To tell if an animal is a ziblet or a flurp, you have to count the buttons and the fingers. Ziblets always have more fingers than buttons.

Let me show you: Here are the buttons of a ziblet and the fingers on one wing. How many buttons does this ziblet have? ... That's right, there are two buttons. And how many fingers? ... That's right, there are three. Are there more fingers or more buttons? Yes, there are more fingers because this is a ziblet. And here are the buttons and the fingers of another ziblet. How many buttons does this ziblet have? ... That's right, there are four. And how many fingers? ... That's right, there are five. Are there more fingers than buttons? Yes, there are more fingers than buttons because this is a ziblet.

Fritz cannot remember this. So we need to remember it for him to help him find a ziblet. Do you know why ziblets have more fingers than buttons? Ziblets have a chemical in their blood that's called zeeken. This chemical makes the fingers of ziblets really sticky, so ziblets can catch insects with their fingers. See, here are the fingers of a ziblet. They are really sticky because ziblets have zeeken in their blood. Ziblets don't need their buttons for anything, so they don't have as many buttons. They always have more fingers than buttons because they catch insects with fingers not with buttons. Fritz doesn't understand this. Can you explain to him? Why do ziblets have more fingers than buttons? (if no or incorrect answer, explain again): Ziblets have more fingers than buttons because they catch food with their fingers not with their buttons. The chemical zeeken makes their fingers sticky so they can catch food with their fingers. Ziblets don't need their buttons for anything, so they don't have so many buttons.

(Experiment 7 only)

Here are the body parts of animals in the pet store. They have a body, antennas, a tail, wings, buttons, and fingers on each wing. And they eat insects that get stuck on their body. To tell if an animal is a ziblet or a flurp, you have to look at the antennas. Ziblets have circles on their antennas and flurps have Xs on their antennas.

Ziblets always have circles on their antennas. Let me show you: Here are the circles on the antennas of a ziblet. And here are the circles on the antennas of another ziblet. This one has circles on its antennas, too, because this is a ziblet.

Fritz cannot remember this. So we need to remember it for him to help him find a ziblet. Do you know why ziblets have circles on their antennas? Ziblets have a chemical in their blood that's called zeeken. This chemical makes the circles on the antennas of ziblets really sticky, so ziblets can catch insects with their antennas. See, here are the circles on the antennas of a ziblet. They are really sticky because ziblets have zeeken in their blood. They always have circles on their antennas because they catch insects with the circles on their antennas. Fritz doesn't understand this. Can you explain to him? Why do Ziblets have circles on their antennas? (if no or incorrect answer, explain again): Ziblets have circles on their antennas

because they catch food with the circles on their antennas. The chemical zeeken makes the circles on their antennas sticky so they can catch food with their circles.

Experiments 2, 3, and 4

This is Mike. He just moved to a new country called Elbee with his family. Everything in Elbee is different from where he used to live. Mike is beginning to learn about all the things in Elbee. His new teacher is helping him.

Mike's new teacher says that things in Elbee sometimes look a lot alike. So she's helping Mike learn how to tell them apart. She showed him some things called ziblets and some things called flurps. She says that ziblets and flurps are two different kinds of things that look a lot alike. Now I'm going to show you some of them.

(Experiments 2 & 3 only)

Here is a ziblet. Can you say ziblet? ... Good! Ziblets have eyes to help them see. See? Here are the eyes. Can you point to the eyes? Good, those are the eyes. Do you know why ziblets have eyes? ... Ziblets have eyes so they can see. Mike doesn't understand this. Can you explain to him? Why do ziblets have eyes? ... (if no or incorrect answer, explain again): Ziblets have eyes so they can see. Ziblets have eyes and eat grapes and live in trees, like this one. What do ziblets eat? That's right, they eat grapes. Where do ziblets live? That's right, they live in trees.

Now, here is a flurp. Can you say flurp? ... Good! Flurps have bolts that help keep them together. See? Here are the bolts. Can you point to the bolts? Good, those are the bolts. Do you know why flurps have bolts? ... Flurps have bolts to keep them together. Mike doesn't understand this. Can you explain to him? Why do flurps have bolts? ... (if no or incorrect answer, explain again): Flurps have bolts to keep them together. Flurps have bolts and use electricity and come from factories, like this one. What do flurps use? That's right, they use electricity. Where do flurps come from? That's right, they come from factories.

So, ziblets and flurps look a lot alike – but they are very different from each other. Ziblets have eyes and eat grapes and live in trees. And flurps have bolts and use electricity and come from factories.

(Experiment 4 only)

Here is a ziblet. Can you say ziblet? ... Good! Ziblets have a round mouth for sucking up grapes. See? Here is a round mouth and [click] here's another round mouth. Can you point to its round mouth? Good, that's its round mouth. Do you know why ziblets have a round mouth? ... Ziblets have a round mouth for sucking up grapes. Mike doesn't understand this. Can you explain to him? Why do ziblets have a round mouth? ... (if no or incorrect answer, explain again): Ziblets have a round mouth for sucking up grapes. Ziblets have a round mouth and eat grapes and live in the jungle, like this. What do ziblets eat? That's right, they eat grapes. Where do ziblets live? That's right, they live in the jungle.

Now, here is a flurp. Can you say flurp? ... Good! Flurps have sharp teeth for crunching up nuts. See? Here are sharp teeth and [click] here are more sharp teeth. Can you point to its sharp teeth? Good, those are its sharp teeth. Do you know why flurps have sharp teeth? ... Flurps have sharp teeth for crunching up nuts. Mike doesn't understand this. Can you explain to him? Why do flurps have sharp teeth? ... (if no or incorrect answer, explain again): Flurps have sharp teeth for crunching up nuts. Flurps have sharp teeth and eat nuts and live in the desert, like this. What do flurps eat? That's right, they eat nuts. Where do flurps live? That's right, they live in the desert.

So, ziblets & flurps look a lot alike – but they are very different from each other. Ziblets have a round mouth and eat grapes and live in the jungle. And flurps have sharp teeth and eat nuts and live in the desert.

Highlights

- We tested two competing accounts for why children use labels in their inductive inferences.
- Participants learned novel categories that varied in their conceptual basis.
- Children based their inductive inferences on conceptual as well as perceptual information.
- Labeled categories are not all alike, and differ in their inductive potential.

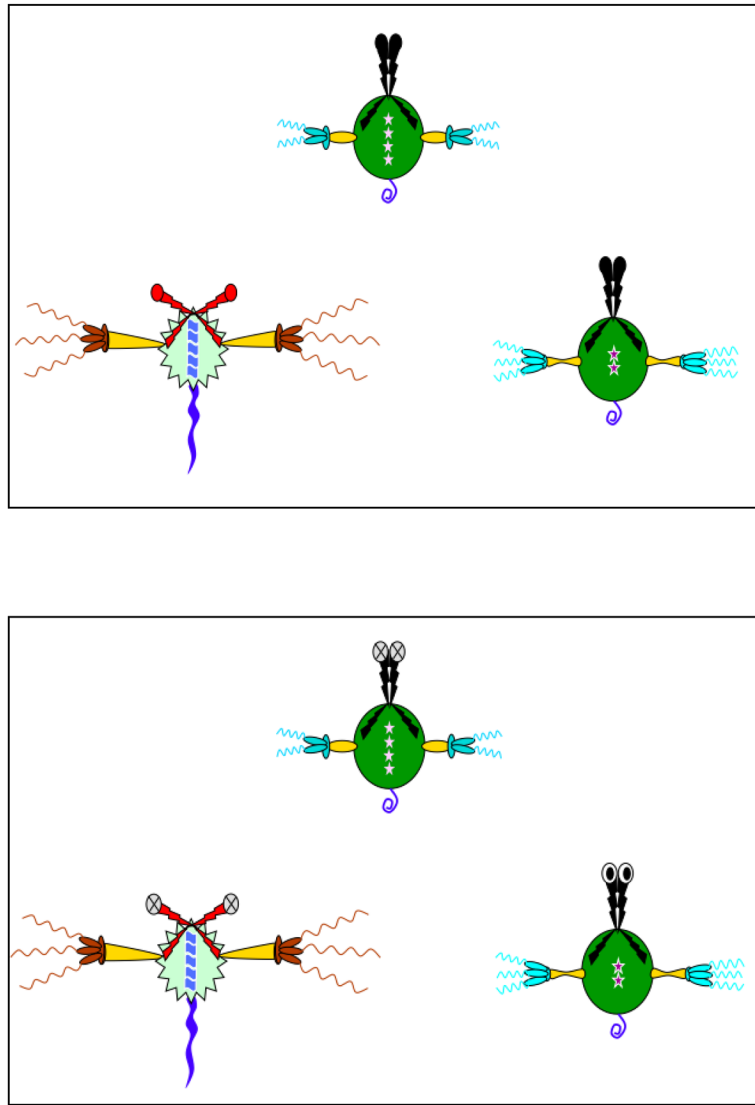


Figure 1. (a) Sample plain picture set (used in SKF and present Experiment 1). (b) Sample modified picture set (used in Experiments 2–7). In each of the two example sets below, the top item and the item on the bottom left are flurps, whereas the item on the bottom right is a ziblet.

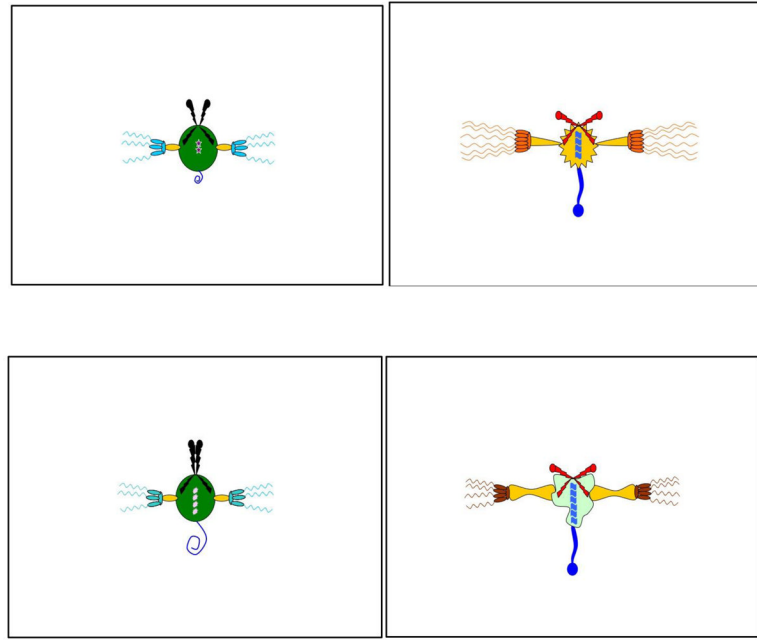


Figure 2. Sample ziblets and flurps (used in SKF and present Experiment 1). The top line displays ziblets (A1C1 and A2C1, left to right, respectively) and the bottom line displays flurps (A1C2 and A2C2, left to right, respectively).

Table 1

Comparison of the experiments.

	Expt. 1	Expt. 2	Expt. 3	Expt. 4	Expt. 5	Expt. 6	Expt. 7
Category Contrast	Non-kind: 2 types of pets	Ontological: animal/artifact	Ontological: animal/artifact	Basic natural kind: 2 kinds of animals	Non-kind: 2 types of pets	Non-kind: 2 types of pets	Non-kind: 2 types of pets
Labels	Ziblet/flurp	Ziblet/flurp	Ziblet/flurp	Ziblet/flurp	Ziblet/flurp	Ziblet/flurp	Ziblet/flurp
Pictures	Plain	Modified	Modified	Modified	Modified	Modified	Modified
Distinguishing Features	Finger:button ratio	Eyes vs. bolts	Eyes vs. bolts	Round mouth vs. sharp teeth	Finger:button ratio	Finger:button ratio	Circles vs. Xs on antennae
Associated Ziblet Features	Nice & friendly; zeeken in blood	Eats grapes; lives in trees	Eats grapes; lives in trees	Eats grapes; lives in jungle	Nice & friendly; zeeken in blood	Nice & friendly; zeeken in blood	Nice & friendly; zeeken in blood
Associated Flurp Features	Wild & dangerous	Uses electricity; from factory	Uses electricity; from factory	Eats nuts; lives in desert	Wild & dangerous	Wild & dangerous	Wild & dangerous
Categorization Task	No picture cue	Habitat picture cues	Habitat picture cues	Habitat picture cues	No picture cue	Rectangle picture cues	No picture cue
Test Properties (with example)	Body parts (round muscles)	Familiar ontol. (wires inside)	Novel (blickets inside)	Novel (blickets inside)	Body parts (round muscles)	Body parts (round muscles)	Body parts (round muscles)

Table 2

Properties used in the induction task, as a function of experiment.

<i>Experiment 1</i>	<i>Experiment 2</i>	<i>Experiments 3 and 4</i>	<i>Experiments 5, 6, and 7</i>
has a big brain*	(Ziblets)	(Ziblets)	(Ziblets)
has a small liver*	breathes	can make a zevy sound	has a big brain*
has a spine in its back	can have babies	has a very sticky toma	has one kidney
has one kidney	has a heart inside	has zimmer inside	has round muscles*
has round muscles*	has a mommy and a daddy	is used for derriping	has three lungs*
has small bones*	has bones inside	needs tiddles to make it move	has tiny ears on the back of its antennae
has soft tissue*	sleeps at night	uses danner	has two hearts*
has thick blood*			
has three lungs*	(Flurps)	(Flurps)	(Flurps)
has tiny ears on the back of its antennae	can be turned off	can help yippets	has a small liver*
has two hearts*	can break	goes outside in the winter	has a spine in its back
has two stomachs	has batteries inside	has a part inside called a cece	has small bones*
	has wires inside	has blickets inside of it	has soft tissue*
	was made by people	has grumpets that make it strong	has thick blood*
	was sold in a store	is good for kertling	has two stomachs

* original properties from SKF

Table 3

Design of Experiments 1–7.

	Task	# Trials	Images per trial	Exptr feedback?
Experimental condition				
• Category training	Training on rule and properties that distinguish ziblets vs. flurps	-		-
• Category learning	Classify instance as ziblet vs. flurp	8	1	yes
• Initial categorization	Classify instance as ziblet vs. flurp	8	1	no
• Induction	Select which test item has same property as target	12	3	no
• Final categorization	Classify instance as ziblet vs. flurp	8	1	no
Control condition				
• Induction	Select which test item has same property as target	12	3	no
• Category training	Training on rule and properties that distinguish ziblets vs. flurps	-		-
• Category learning	Classify instance as ziblet vs. flurp	8	1	yes
• Initial categorization	Classify instance as ziblet vs. flurp	8	1	no

Table 4

Mean number of category-based inductive inferences in the experimental conditions (out of 12; chance is 6). SDs in parentheses.

Expt	Description	Child Experimental	Child Control	Child E>C?	Adult Experimental	Adult Control	Adult E>C?
1	Slousky, Kloos, Fisher (SKF) replication	3.31* (1.74)	---	N/A	9.00* (3.73)	---	N/A
2	Ontological distinction, familiar properties	9.37* (3.28)	5.50 (2.68)	yes	11.94* (0.25)	3.81* (2.49)	yes
3	Ontological distinction, novel properties	8.13* (3.86)	4.88 [±] (2.31)	yes	8.81* (3.90)	2.63* (1.86)	yes
4	Basic-level distinction, novel properties	7.88 [±] (3.91)	5.00 (3.27)	yes	9.06* (3.77)	3.25* (1.91)	yes
5	SKF categories with Expts 2–4 stim pix	3.25* (2.62)	5.00 (2.90)	no	7.69 (4.70)	3.56*	yes
6	SKF categories with Expts 2–4 stim pix and procedure	4.00* (2.56)	4.13* (1.71)	no	---	---	---
7	SKF categories with Expts 2–4 stim pix & disting. features	5.56 (4.13)	---	no ^d	---	---	---

^dThis comparison involves the control conditions of Experiments 5 and 6.

* indicates significantly different from chance ($p < .05$);

[±] indicates a non-significant trend ($p < .10$).