Changes in Beliefs About Category Homogeneity and Variability Across Childhood

Amanda C. Brandone
Lehigh University

Effective category-based induction requires understanding that categories include both fundamental similarities between members and important variation. This article explores 4- to 11-year-olds’ ($n = 207$) and adults’ ($n = 49$) intuitions about this balance between within-category homogeneity and variability using a novel induction task in which participants predict the distribution of a property among members of a novel category. Across childhood, children learned to recognize variability within categories—showing increasing sensitivity to the role of property type and domain in constraining inferences. Children below the age of 6 showed evidence for a domain-general assumption that categories are homogeneous—generalizing properties broadly to 100% of category members. These studies support important developmental changes in category representations that may influence category-based induction.

From early in development, children use categories to guide their learning and interactions with the environment. Categories are an important conceptual tool because they allow us to make broad inferences about the world based on limited evidence. For example, if one learns that an individual frog grew from a tadpole, has a sticky tongue, and is cold-blooded, one can reasonably generalize these facts beyond that specific exemplar to other instances of the category “frog.” This process relies on an assumption of category homogeneity—the idea that categories mark groups of individuals that are highly similar to one another (Gelman, 2003; Murphy, 2002). It is also the case, however, that the members of categories are not homogeneous with respect to all properties. For example, individual frogs may differ considerably in color, habitat, toxicity, and whether or not they are nocturnal. They may also differ on various transient properties including being wet, hungry, dirty, or sick. Thus, categories encompass a balance between shared characteristics that are common across category members and distinct characteristics of individuals. This tension between similarity and variability within categories is a key component of our category knowledge and is required in order to effectively use categories as a basis for induction.

The present research aims to explore the development of children’s intuitions about this balance between within-category homogeneity and variability. Specifically, using a novel induction task, two studies examine the extent to which children between the ages of 4 and 11 assume that members of novel categories are fundamentally similar and when they can appropriately constrain these assumptions to recognize variability within categories.

There is extensive evidence from both adults and children documenting the role of categories in induction (e.g., Carey, 1985; Coley, Hayes, Lawson, & Moloney, 2004; Gelman & Markman, 1986; Heit, 2000; Osherson, Smith, Wilkie, Lópex, & Shafir, 1990; Rips, 1975). Studies show that when learning a new fact about an instance of a category, young children (and even infants; Graham, Kilbreath, & Welder, 2004) apply knowledge about category membership to guide their inferences, readily generalizing information from one member of a category to another, even when the instances differ perceptually (Gelman & Markman, 1986; see Sloutsky, Kloos, & Fisher, 2007 for an alternative argument). This ability is important given that much of human reasoning and conceptual development involves making category-based inferences that extend beyond the available evidence. However, as
noted previously, there is also a need for constraints on this induction process: Not all information about individuals is appropriately generalized on the basis of category membership (see Coley, 2012 for a similar argument).

As adults, we have a rich store of conceptual knowledge that can guide us in deciding when to expect a property of an individual category member to be widely shared and thus generalize, and when to expect within-category variability and thus limit our generalizations. One powerful piece of this knowledge is information about the nature of the property and the extent to which it is considered "projectable" (Goodman, 1983). Adults rely on statistical knowledge about the frequency of properties to guide their generalizations and have been shown to exercise more caution in generalizing when the property is thought to be infrequent or heterogeneous within the category (Nisbett, Krantz, Jepson, & Kunda, 1983). Likewise, adults view biological, stable, and "essential" (Gelman, 2003) properties (e.g., parts, function, internal structure) as having more inductive power than more superficial, transient, or idiosyncratic properties (e.g., broken legs, wet fur; Cimpian, Brandone, & Gelman, 2010). Knowledge about the nature of the target category, including its ontological domain, also influences patterns of induction because naturally existing categories like animals and plants are viewed as more coherent and richly structured than artifact categories (Coley et al., 2004; Gelman, 1988; Shipley, 1993). Finally, additional knowledge about the similarity of the individual and conclusion category, the typicality of the category, and the coherence, specificity, taxonomic level, and informativeness of the conclusion category can also play a role in guiding and constraining adults' induction (e.g., Goodman, 1983; Heit, 2000; Nisbett et al., 1983; Osherson et al., 1990; Rips, 1975).

The question I examine here is the extent to which young children share this knowledge about the constraints on induction. There is some evidence to suggest that young children may be biased to assume that categories are largely homogeneous. For example, preschoolers readily draw broad, category-wide inferences after being exposed to only a single member of a novel category (Brandone & Gelman, 2009, 2013). Preschoolers have also been shown to be highly motivated to acquire information about categories—preferring learning facts and explanations about categories to learning analogous information about individuals (Cimpian, 2016). Young children also rely on category knowledge over individuating information when making inferences about the future behavior of individual category members (Berndt & Heller, 1986; Taylor, 1996). Finally, evidence suggests that young children ignore important within-category variability in induction tasks where they are asked to select between or evaluate the inductive strength of diverse versus nondiverse samples (e.g., two Dalmatians vs. a Dalmatian and a poodle; Gutheil & Gelman, 1997; Rhodes & Brickman, 2010; Rhodes & Liebenson, 2015). These data lend support to the view that children show an early-emerging bias to construe categories as largely homogeneous and thus overlook the variability that exists within categories.

However, there is also some evidence to suggest that children are sensitive to some of the same factors that constrain induction in adults. For example, studies of generic language use show that even preschoolers express more broad, category-wide generalizations when talking about categories of fictional animals as compared to artifacts, suggesting that they may be sensitive to the influence of category domain on homogeneity (Brandone & Gelman, 2009, 2013). Classic induction tasks that test whether children extend a property from one category member to a new exemplar also indicate awareness that some properties and some categories are more suitable for inductive reasoning than others. For example, as early as preschool, children reason differently at different levels of the taxonomic hierarchy—making more projections within the more coherent, subordinate level (e.g., a daffodil to another) than at higher levels (e.g., a daffodil to a house plant; Gelman, 1988; Gelman & O’Reilly, 1988). Preschoolers also limit their inferences based on the nature of the to-be-generalized property—showing particular sensitivity to whether the property is stable and internal or more transient and idiosyncratic (Gelman, 1988; Graham, Welder, & McCremon, 2003). By second grade (but not before), children distinguish among categories in the inferences they draw—appropriately drawing more inferences within natural kind categories (i.e., animals, plants) than within artifact categories (Gelman, 1988; Gelman & O’Reilly, 1988). Finally, between 6 and 10 years, children utilize domain-relevant contextual and causal knowledge to systematically guide their inferences (e.g., extending diseases on the basis of habitat, but internal parts on the basis of taxonomic relations; Coley, 2012). Thus, collectively, the existing research indicates that, although young children may have an early-emerging bias to assume that categories are homogeneous, children are also acquiring knowledge
about how to selectively limit induction from early in development.

Despite these significant contributions, there are limitations to the existing literature. First, although children’s generic language use provides a compelling initial estimate of children’s beliefs about category coherence (as generics express generalizations and are uniquely suited to talking about homogeneous categories), generics provide only an indirect measure of these beliefs. Likewise, although classic category-based induction tasks tell us whether children generalize information from one category member to another (e.g., generalizing “likes to eat alfalfa” from a gray rabbit to a brown rabbit; Gelman, 1988), these data provide only a partial picture of children’s beliefs about homogeneity versus variability within categories. Consider the conclusions that can be drawn from a standard induction measure in which a child extends a property to a new category member. The child could generalize because she believes the property to be present universally—in 100% of category members; however, this response is also appropriate for a property that exists in the vast majority or even just over half of category members. Likewise, a child who limits generalization may do so based on a belief that the property is present in a handful of category members, a substantial minority, or just the exemplar member itself. Thus, although standard category-based induction measures are informative, there are limits to the conclusions that can be drawn from them. In order to fully assess children’s beliefs about category structure, data on how children think about the distribution of properties within the larger category are needed. That is, we need data revealing not just whether children generalize a property (e.g., eating alfalfa) from one category member to the next (e.g., from a gray rabbit to a brown rabbit), but also how children believe that property is distributed among the entire category (e.g., in the larger category “rabbits”). No existing studies have provided this in-depth account of children’s category expectations.

Given this context, the current studies aim to examine three key questions. First, how do children think about the distribution of properties among category members? That is, given a novel category, to what extent do children believe that category to be homogeneous? Second, how do children’s expectations about category homogeneity differ depending on factors shown to influence adults’ judgments? Specifically, do children recognize the role of the nature of the target property and the ontological domain of the category in limiting generalizations? Finally, how do children’s assumptions about within-category homogeneity versus variability change across childhood? In particular, with age, do children refine their views about the balance between homogeneity and variability within categories?

Two studies were conducted to address these questions. Both involved a novel induction task in which children (ages 4–11) were introduced to a series of novel categories on an alien planet (i.e., “Droid’s planet”). Children were shown a member of each category (e.g., “This is a kind of animal called a floom”), told about a property of that category member (e.g., “This floom is orange”), prompted to think about the entire category (e.g., “Now think about all the flooms on Droid’s planet”), and asked to provide an estimate of the prevalence of the property within the larger category. In Study 1, prevalence estimates were given using a 4-point scale (100%, 80%, 50%, 20%). In Study 2, prevalence expectations were indicated through forced choice questions about whether the property was present in “all,” “lots,” or “just a few” category members.

Children’s sensitivity to the role of category domain in limiting induction was examined by assessing their inferences about items from three domains: animal, artifact, and social categories. The animal and artifact domains were selected because they have been used frequently in prior research examining domain differences (Opfer & Gelman, 2010) and because they represent distinct and distant points on the continuum from natural and essentialized to human-made and conventionalized (Diesendruck & Gelman, 1999; Kalish, 1995). Social categories were selected as a comparison because they share features with both naturally occurring and constructed categories (Haslam, Rothschild, & Ernst, 2000; Hirschfeld, 1998; Rhodes & Gelman, 2009), are variable in the extent to which they are viewed as essentialized (e.g., gender) versus conventionalized or ad hoc (e.g., Midwesterners, movie buff; Haslam et al., 2000), and because the question of how children think about variability has unique implications in the social domain (Allport, 1954; Macrae & Bodenhausen, 2000). The use of property type in constraining induction was also examined by assessing children’s responses to four property types: parts (e.g., has ears, has a handle), behavior or function (e.g., slides on its belly, is used to reach things), color (e.g., is orange, red), and temporary or accidental properties (e.g., is dirty, has a broken handle). These property types were selected to vary in the extent to which they are
viewed as generalizable across category members, ranging from highly generalizable in the case of parts to quite limited in the case of temporary or accidental properties.

Consistent with prior research suggesting a tendency to construe basic-level categories as largely homogeneous in early childhood (e.g., Brandone & Gelman, 2009, 2013; Gelman, 2003; Rhodes & Brickman, 2010), I predicted that young children would assume that members of the novel test categories are fundamentally similar to one another and thus show a broad preference for the 100% or “all” prevalence choices—across domain and property type. Consistent with evidence of increasing knowledge about property types and domains across childhood (e.g., Gelman, 1988; Gelman & O’Reilly, 1988), I predicted that the homogeneity bias would diminish as children acquire conceptual knowledge that helps them to recognize variability within categories and thus appropriately constrain their inferences. Specifically, I predicted that with age children would increasingly vary their inferences based on domain and property type.

Study 1

Method

Participants

Participants included thirty 4-year-olds (19 boys, 11 girls; $M = 4.49$ years, $SD = 0.27$), thirty 5- and 6-year-olds (18 boys, 12 girls; $M = 5.64$, $SD = 0.45$), twenty-five 7- and 8-year-olds (13 boys, 12 girls; $M = 7.86$, $SD = 0.58$), twenty-six 9-, 10-, and 11-year-olds (10 boys, 16 girls; $M = 10.14$, $SD = 0.78$), and twenty-four undergraduates (12 men, 12 women; $M = 19.44$, $SD = 1.19$). Children were recruited from schools and a database of families interested in developmental research from communities in and around a midsize northeastern U.S. city. Although demographic data were not collected on the full sample, census data indicate that residents of this area are primarily White (79%) with 3% of residents identified as Asian, 6% as Black or African American, and 12% as Hispanic or Latino. Undergraduates were recruited from an Introduction to Psychology participant pool at a small private university in the same city and participated for course credit. Forty-four percent of undergraduate participants identified as White, 4% as Black or African American, 32% as East or Southeast Asian, 4% as Hispanic, and 16% as Middle Eastern. Data collection took place from June 2013 to July 2014.

Stimuli

Color illustrations of a single exemplar from 24 novel categories were created. Categories included eight items from each of three domains: animal, artifact, and social categories. Each exemplar displayed a target property from a set of four property types: parts (e.g., has ears, a handle, arms), behavior or function (e.g., slides on its belly, is used to reach things, eats flowers), color (e.g., is orange, red, green), and temporary or accidental properties (e.g., is dirty, has a broken handle, has a broken leg). Property types were distributed equally across domains such that each included two items targeting each property type. The study thus used a 3 (domain: animal, artifact, social) $\times$ 4 (property type: parts, behavior or function, color, temporary or accidental) $\times$ 2 (item) design (see Appendix for the full set of items and properties).

Procedure

Introduction Phase

Children were tested individually in a quiet room with an experimenter. Children were first introduced to an alien puppet named Droid and his alien planet where there are all different kinds of new animals, tools, and people. Children were told that they were going to play a game in which they answered questions about the animals, tools, and people on Droid’s planet. (Given that children were unlikely to be familiar with the term “artifact,” the superordinate category label “tool” was used to describe all of the artifact stimuli.)

Children were then introduced to the response scale. The scale was created using pictures of four samples of dots in which 100%, 80%, 50%, and 20% were filled in (see Figure 1). These percentages were selected to represent a large range of possible responses and to display clearly distinguishable prevalence levels. Points on the scale were labeled initially with quantitative terms (e.g., all, lots, some, just a few) to assist children in how to interpret the proportions in each sample. However, the response scale samples were labeled only during the introduction and practice trials; they were not labeled during the primary task (see below). Children were instructed in how to use the scale and given a chance to point to each of the four samples.
Practice Trials

After children were familiarized with the response scale, they completed two sets of practice trials. In the first, children were presented with four sets of shapes each representing one of the four answer choices. For example, children were shown a set of 22 squares, 17 of which were blue, and told to look at the picture and decide, “How many of the squares are blue: All of them? Lots of them? Some of them? Or just a few?” As the experimenter read each answer choice she pointed to the corresponding sample on the scale. Children responded by pointing. These practice items were intended to help children understand the relation between the target sample (e.g., 17 of the 22 squares) and the samples of dots on the response scale (e.g., 80%). Incorrect responses were corrected, and these items were repeated until children responded correctly for each one. Children’s initial responses were correct on 78.2% of items (4-year-olds: 59.7%; 5- and 6-year-olds: 79.8%; 7- and 8-year-olds: 91.0%; 9-, 10-, and 11-year-olds: 96.2%).

The second set of practice items more closely approximated the test stimuli. Children were shown a single exemplar from each of four fruit categories (e.g., red strawberry, brown banana, green grapes, and red apple). For each item, the color of the exemplar was identified (e.g., “This apple is red”), and children were asked to think about the entire category (e.g., “Think about all the apples in the world”) and estimate the prevalence of the property in that category (e.g., “How many of them do you think are red?”). After reading the question, the experimenter pointed to each answer choice on the response scale (without labeling it) to invite children to respond. Children responded by pointing.

Primary Task

Following the practice items, the primary task began. Children were presented with the novel stimuli one at a time on a tablet computer. For each item, children heard a sequence such as the following:

This is a kind of animal [tool, person] called a floom. There are lots of flooms on Droid’s planet. This floom is orange [blue, has green skin]. Now think about all the flooms on Droid’s planet. How many of them do you think are orange [blue, have green skin]?

After reading the question, the experimenter pointed to each answer choice on the response scale (without labeling it) to invite children to respond. Children responded by pointing.

Items from the same domain were presented in blocks and the domain was specified before each block. Children were shown a set of familiar items from the domain (e.g., for tools, the set included a hammer, shovel, scissors, screwdriver, spatula, etc.) and were told, “These are all different kinds of tools on our planet. Now we’re going to see some of the tools on Droid’s planet.” This was done to
highlight the domain for each block and to help anchor children’s responses in their knowledge of that domain. This was particularly important for the artifact domain due to uncertainty regarding the youngest children’s familiarity with the label “tools.” The order of the blocks was counterbalanced across participants. Within each domain, four random orders of items were created and item orders were counterbalanced across participants.

**Modifications for Adults**

The procedure for adults was comparable to that for children with some modifications. Adults were tested using a written format such that they read the information to themselves and circled the appropriate sample of dots on the response scale. Counterbalancing, stimuli, item labels, questions, and response prompts were identical to those given to children except that the alien puppet and his planet were not referenced and adults were asked to mark their answer on a response sheet. The study was introduced by explaining that participants were going to learn about a new planet that scientists recently discovered in a remote galaxy. As with the children, the adults completed two sets of practice items and then moved to the primary task. Adults were not corrected for incorrect practice item responses.

**Results**

The goal of Study 1 was to examine how children think about the distribution of properties among novel category members and to what extent they differentiate their inferences based on category domain (animal, artifact, social) and property type (part, behavior or function, color, temporary or accidental). To address these questions, I converted participants’ prevalence level selections to a numerical 1–4 scale (100% selections = 4, 80% selections = 3, 50% selections = 2, and 20% selections = 1) and entered them into a 3 (domain: animal, artifact, social) × 4 (property type: within) × 5 (age group: between) analysis of variance. Results revealed main effects of property type, $F(3, 390) = 117.97$, $p < .001$, $\eta^2_p = .48$, and category domain, $F(2, 260) = 3.08$, $p = .048$, $\eta^2_p = .02$, as well as two-way interactions of Property Type × Domain, $F(6, 780) = 7.28$, $p < .001$, $\eta^2_p = .05$, and Property Type × Age Group, $F(12, 390) = 16.78$, $p < .001$, $\eta^2_p = .34$. To disentangle these interactions, I examined the effects of property type and domain separately for each age group and conducted follow-up pairwise comparisons with sequential Bonferroni corrections (see Figure 2).

Analyses revealed that adults clearly varied their responses by property type and domain. Results showed main effects of property type, $F(3, 69) = 115.11$, $p < .001$, $\eta^2_p = .83$, and domain, $F(2, 46) = 9.15$, $p < .001$, $\eta^2_p = .29$, as well as a Property Type × Domain interaction, $F(6, 138) = 5.85$, $p < .001$, $\eta^2_p = .20$. As can be seen in Figure 2a, adults made broad generalizations about category member parts and, to a somewhat lesser extent, their behavior or function ($p < .001$). They made more limited generalizations about color and even fewer about temporary or accidental properties (all $ps < .001$). Adults also acknowledged differences in homogeneity based on domain, extending properties more broadly overall for animals than for artifacts and social categories (both $ps < .05$). Finally, adults showed nuanced differentiation by domain within each property type. For example, adults acknowledged less homogeneity in the color of artifacts relative to the color of animals and social categories (both $ps < .05$). Likewise, they recognized more homogeneity in the parts and less homogeneity in the behavior of social categories relative to animals and artifacts (all $ps < .05$). Overall, these findings reveal clear differences in adults’ homogeneity expectations based on category domain and property type.

The oldest group of children (9- to 11-year-olds) showed a very similar pattern of responses. Analyses revealed a main effect of property type, $F(3, 75) = 33.86$, $p < .001$, $\eta^2_p = .58$, and an interaction of Property Type × Domain, $F(6, 150) = 3.16$, $p = .006$, $\eta^2_p = .11$. The main effect of domain was nonsignificant ($p = .43$). 

As can be seen in Figure 2b, 9- to 11-year-olds made broad generalizations about parts and, to a somewhat lesser extent, behavior or function ($p < .05$). Also like adults, they made more limited generalizations about color and restricted their generalizations further for temporary or accidental properties (all $ps < .001$). Although 9- to 11-year-olds did not show overall domain differences in their responses ($ps > .20$), they did show some differences in inferences by domain and property type. Most notably, they limited their homogeneity estimates for behavior within social categories relative to animal and artifact categories (both $ps < .05$). Overall, findings suggest that 9- to 11-year-olds, like adults, use their knowledge about domain and property type to limit their category-based inferences.

The next youngest group of children, 7- to 8-year-olds, showed a further attenuated pattern.
Analyses revealed a main effect of property type only, $F(3, 72) = 33.57, p < .001$, $\eta^2_p = .58$. The main effect of domain and the interaction of Property Type $\times$ Domain were nonsignificant ($ps > .12$). As can be seen in Figure 2c, 7- to 8-year-olds made broad generalizations about parts and, to a somewhat lesser extent, behavior or function ($p < .05$). They also recognized less homogeneity in color ($ps < .01$) and even less for temporary and accidental properties ($ps < .001$). However, 7- to 8-year-olds showed little differentiation based on domain ($ps > .55$), with the exception of generalizing the parts of people more broadly than the parts of animals or artifacts (both $ps < .05$). Overall, data suggest that by 7–8 years, children are able to use some basic, domain-general knowledge about property type to constrain their inferences.

Finally, due to the similarity of responses for the 4-year-olds and 5- to 6-year-olds, I report the results of these age groups combined. Analyses revealed a
main effect of property type only, $F(3, 177) = 2.91$, $p = .036$, $\eta^2_p = .05$. The main effect of domain and the interaction of Property Type $\times$ Domain were nonsignificant ($ps > .32$). Despite the main effect of property type, the youngest children did not resemble older children and adults in their responses and showed very little differentiation in their homogeneity expectations (see Figure 2d). Most notably, 4- to 6-year-olds failed to acknowledge less homogeneity for temporary or accidental properties as compared to more stable and kind-relevant properties including parts, behavior, and color ($ps > .30$). These data demonstrate that 4- to 6-year-olds fail to make use of any knowledge they may have about domain and property type to guide their category-based inferences.

Although overall analyses did not reveal a main effect of age group in participants’ responses, $F(4, 130) = 0.61$, $p = .66$, in order to specifically examine whether young children show evidence of a bias to expect high levels of similarity across members of novel categories, I examined the frequency with which participants selected the 100% sample when generalizing properties of the target stimuli. A series of one-sample $t$-tests was conducted to examine whether the percentage of 100% choices was greater than would be expected by chance (25%) within each age group. Results revealed a small but significant preference for the 100% sample in the youngest participants—4-, 5- and 6-year-olds, $t(59) = 3.09$, $p = .003$, but not in 7- and 8-year-olds, 9- to 11-year-olds, or adults ($ps > .17$; see Figure 3). See Table 1 for the percentage of 100%, 80%, 50%, and 20% selections by age group, domain, and property type. Overall, these results are consistent with the hypothesis that, when learning about novel categories, young children show a modest tendency to extend properties of category members to 100% of the category.

**Discussion**

The results of Study 1 point to two main conclusions. First, children’s sensitivity to the ways in which different kinds of properties and categories limit generalizability increased with age. Specifically, although results offered little support for preschoolers’ awareness of the role of these factors, data demonstrated sensitivity to the role of property type in limiting induction by ages 7–8 and analogous sensitivity to the role of category domain by ages 9–11. Second, consistent with suggestions in prior work (e.g., Gelman, 2003; Rhodes & Brickman, 2010), young children showed a modest domain-general bias to assume that categories are homogeneous. This bias, observed in a small but reliable tendency to generalize properties of individual category members to 100% of the category, became less pronounced with age.

In interpreting 4- to 6-year-olds’ homogeneity bias and failure to limit generalizations by property type and domain, it is important to consider the complexity of the induction task used here. To be successful, children needed to imagine the existence of a novel planet with novel categories, consider the potential variability among members of those categories, recruit and apply any conceptual knowledge they may have about domains and property types, and then use an abstract scale to communicate their responses. Difficulty could stem from problems at each level. Thus, although preschoolers’ preference for the 100% sample and lack of differentiation in responses by domain and property type could reflect an early, domain-general bias to construe categories as largely homogeneous, it could also reflect a strategy on the part of the youngest children to select the first answer choice on the scale in the face of this complex task. Study 2 was designed to help address this possibility. The
**Table 1**
Study 1: Percentage of Selections by Age, Domain, Prevalence Level, and Property Type

<table>
<thead>
<tr>
<th>Age group</th>
<th>Domain</th>
<th>Prevalence level %</th>
<th>Percentage of selections</th>
<th>Prevalence level %</th>
<th>Percentage of selections</th>
<th>Prevalence level %</th>
<th>Percentage of selections</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Parts</td>
<td>Behavior/ function</td>
<td>Color</td>
<td>Temporary/ accidental</td>
<td>Parts</td>
</tr>
<tr>
<td>Adults</td>
<td>Animal</td>
<td>100</td>
<td>58.3**</td>
<td>39.6*</td>
<td>10.4</td>
<td>4.3</td>
<td>25.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>80</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>50</td>
<td>12.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>20</td>
<td>4.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Artifact</td>
<td>100</td>
<td>35.4</td>
<td>43.8*</td>
<td>0.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>80</td>
<td>43.8*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>50</td>
<td>18.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>20</td>
<td>2.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Social</td>
<td>100</td>
<td>75.0**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>80</td>
<td>18.8</td>
<td>35.4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>50</td>
<td>6.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>20</td>
<td>0.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9–11 years</td>
<td>Animal</td>
<td>100</td>
<td>50.0**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>80</td>
<td>21.2</td>
<td>28.8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>50</td>
<td>21.2</td>
<td>23.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>20</td>
<td>7.7</td>
<td>9.6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Artifact</td>
<td>100</td>
<td>40.4*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>80</td>
<td>23.1</td>
<td>30.8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>50</td>
<td>25.0</td>
<td>23.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>20</td>
<td>11.5</td>
<td>5.8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Social</td>
<td>100</td>
<td>55.8**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>80</td>
<td>17.3</td>
<td>23.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>50</td>
<td>21.2</td>
<td>36.5*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>20</td>
<td>5.8</td>
<td>19.2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7–8 years</td>
<td>Animal</td>
<td>100</td>
<td>44.0*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>80</td>
<td>22.0</td>
<td>26.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>50</td>
<td>22.0</td>
<td>30.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>20</td>
<td>12.0</td>
<td>10.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Artifact</td>
<td>100</td>
<td>46.0**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>80</td>
<td>22.0</td>
<td>20.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>50</td>
<td>10.0</td>
<td>18.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>20</td>
<td>22.0</td>
<td>16.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Social</td>
<td>100</td>
<td>66.0**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>80</td>
<td>14.0</td>
<td>26.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>50</td>
<td>16.0</td>
<td>26.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>20</td>
<td>4.0</td>
<td>18.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4–6 years</td>
<td>Animal</td>
<td>100</td>
<td>37.5**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>80</td>
<td>20.8</td>
<td>19.2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>50</td>
<td>18.3</td>
<td>19.2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>20</td>
<td>23.3</td>
<td>22.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Artifact</td>
<td>100</td>
<td>35.0*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>80</td>
<td>18.3</td>
<td>23.3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>50</td>
<td>28.3</td>
<td>19.2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>20</td>
<td>18.3</td>
<td>22.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Social</td>
<td>100</td>
<td>43.3**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>80</td>
<td>20.8</td>
<td>18.3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>50</td>
<td>22.5</td>
<td>19.2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>20</td>
<td>13.3</td>
<td>30.8</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Note.** Bolded values indicate the most frequent response for a given item type within each age group. Comparisons against chance values of 25%: *p < .10. *p < .05. **p < .01.
goal of Study 2 was to replicate Study 1 using a simplified task in order to potentially reveal greater sensitivity among younger children.

**Study 2**

Study 2 examined how children think about the distribution of properties among category members using a simplified response scale. I again compared participants’ inductive inferences across three domains and four property types; however, instead of asking children to report their prevalence estimates using a complex dots scale, I used a series of forced choice questions (e.g., “Do you think all of them are orange?”; “Do you think lots of them are orange?”; “Do you think just a few of them are orange?”). If young children are indeed sensitive to the role of domain and property type in limiting category homogeneity, this forced choice task should better enable them to demonstrate this knowledge. Given the similarity of responding between adults and 9-to 11-year-olds in Study 1, Study 2 focused on younger participants—examining separate groups of 4-year-olds, 5-year-olds, 6-year-olds, and 7- to 8-year-olds, as well as an adult comparison group.

**Method**

**Participants**

Participants included twenty-four 4-year-olds (12 boys, 12 girls; $M = 4.61$ years, $SD = 0.29$), twenty-five 5-year-olds (12 boys, 13 girls; $M = 5.52$, $SD = 0.28$), twenty-four 6-year-olds (10 boys, 14 girls; $M = 6.44$, $SD = 0.29$), twenty-three 7- and 8-year-olds (9 boys, 14 girls; $M = 7.84$, $SD = 0.66$), and 25 adults (13 boys, 12 girls; $M = 31.92$, $SD = 8.73$). Children were recruited as in Study 1. Adults were recruited through Amazon’s Mechanical Turk, an online crowd-sourcing platform, and were compensated $.40 for participation. Adults came from a relatively diverse racial and ethnic background: 52% identified themselves as Asian, 40% as White or Caucasian, and 8% as Hispanic or Latino. Data collection took place between June 2014 and March 2015.

**Stimuli**

The stimuli were identical to Study 1 (see Appendix).

**Procedure**

**Introduction Phase**

As in Study 1, children were first introduced to the alien named Droid and told that they were going to play a game in which they answered questions about the animals, tools, and people on Droid’s planet. Children were then shown a series of three samples of circles (one at a time) in which 100%, 75%, and 17% were colored green. The experimenter described each sample using a quantified statement (e.g., “Look at this picture. All the circles are green, right? Now look at this picture. All of the circles are not green, right? But lots of them are green. Now look at this picture. All of the circles are not green, right? Just a few of these circles are green.”). The purpose of this exercise was to introduce the quantifiers used in the central task and to foreshadow the all versus not all considerations children were asked to make later.

The first set of practice trials tested children’s ability to answer yes or no questions about each quantified statement. Children were presented with three sets of shapes each representing one of the three answer choices. Children were first shown a set of 22 blue squares and asked “Are all the squares blue?” Children were then shown a set of 18 triangles, 14 of which were orange and asked “Are all the triangles orange?” and, subsequently, “Are lots of the triangles orange?” Finally, children were shown a set of 23 stars of which 3 were pink and asked, “Are all the stars pink?,” “Are lots of the stars pink?,” and “Are just a few of the stars pink?” Children responded by answering “yes” or “no.” Incorrect responses were corrected and the items were repeated until children responded correctly for each. Initial responses were correct on 97.3% of Set 1 practice items (4-year-olds: 93.1%; 5-year-olds: 95.8%; 6-, 7-, and 8-year-olds: 100%).

The second set of practice items more closely resembled the test stimuli. Children were shown a single exemplar from each of three fruit categories (e.g., green apple, brown banana, and red strawberry). The color of the exemplar was identified (e.g., “This apple is green”), and children were asked whether that color characterized all category members (e.g., “Do you think all apples green?”). If children answered “yes,” the experimenter moved to the next item. If they responded “no,” the experimenter asked whether that color characterized lots of category members (e.g., “Do you think lots of apples are green?”). If children again responded “no,” the experimenter asked whether that property characterized just a few category members (e.g., “Do
you think *just a few* apples are green?”). Incorrect responses were not corrected and children responded correctly on 72.7% of items (4-year-olds: 63.9%; 5-year-olds: 72.2%; 6-year-olds: 73.6%; 7- and 8-year-olds: 79.2%). Although performance on these practice items was far from perfect, this is in part due to the expected ambiguity of these questions. For example, although it is clear that not all apples are green (confirmed by 98.9% of children), it is less clear whether one should respond “yes” to the question, “Do you think *lots* of apples are green?” (as children “correctly” did 67.4% of the time). Likewise, although it is clear that not all bananas are brown (confirmed by 95.8% of children), it is less clear whether one should respond “yes” to the question “Do you think *lots* of bananas are brown?” (as children “incorrectly” did 30.5% of the time). Using less stringent criteria, only 2.5% of responses were considered totally incorrect (i.e., “yes” to all apples are green, all bananas are brown, just a few strawberries are red; 4-year-olds: 6.9%; 5-year-olds: 1.4%; 6-year-olds: 1.3%; 7- and 8-year-olds: 0%).

**Primary Task**

In the primary task, children were presented with the novel stimuli one at a time and for each heard a sequence such as the following:

“This is a kind of animal [tool, person] called a *foom*. There are lots of *fooms* on Droid’s planet. This *foom* is orange [is blue, has green skin]. **Now think about all the *fooms* on Droid’s planet.**

The researcher then asked whether the target property was present in all category members (e.g., “Do you think *all* of them are orange?”). If children answered “yes,” the experimenter moved to the next item. If children responded “no,” the experimenter asked whether the property characterized *lots* of category members (e.g., “Do you think *lots* of them are orange?”). Following “yes” responses, the experimenter moved to the next item. Following “no” responses, the experimenter asked whether the property characterized *just a few* category members (e.g., “Do you think *just a few* of them are orange?”). (Note that given this forced choice response mode, it was possible for children to respond “no” to all of the questions. Eight children responded this way on at least one item—often explaining their choice by indicating that the property was present in the pictured exemplar only. In order to include these responses [24 of 1,152 total] in analyses, I treated these cases as equivalent to “yes” responses to the *just a few* question. Importantly, there is no difference in the pattern of results if these items are dropped from the analyses.)

As in Study 1, items from the same domain were presented in blocks. The target domain was specified before each block, and children were shown a set of familiar exemplars from the domain. The order of the blocks was counterbalanced across participants. Within each domain, four random item orders were created, and these orders were counterbalanced across participants.

**Modifications for Adults**

The adults were tested using Qualtrics software (Provo, UT) through the Mechanical Turk platform. The task was again introduced by explaining that participants were going to learn about the animals, tools, and people on a remote planet. The alien was not referenced and there were no practice trials. For each item, a picture appeared with the target information (e.g., “This is a kind of animal called a *foom*. There are lots of *fooms* on the new planet. This *foom* is orange.”) Participants were then prompted to consider the entire category (e.g., “**Now think about all the *fooms* on the new planet.**”) and evaluate the prevalence of the target property within that category. The format of this prompt differed for adults in that instead of answering sequential yes or no questions regarding all, lots, and just a few category members, adults selected their answer from among three simultaneously presented options. Items from the same domain were presented in blocks and the order of the block presentation was randomized. Within each domain, the order of the items was also randomized.

**Results**

The goal of Study 2 was to again examine how children think about the distribution of properties among category members and to what extent they constrain their inferences based on category domain and property type. However, the methodological changes made in Study 2 require a different analytic strategy than in Study 1. Because of the difference in the response format (i.e., yes or no responses to forced choice questions), Study 2 data were analyzed by examining the likelihood of providing a “yes” response to the test questions. In addition, given that whether children received the lots or just a few questions was contingent on their response to the all question, separate analyses were
conducted for each question (all, lots, and just a few). These analyses used the generalized estimating equations (GEE) procedure because it can assess both within- and between-subject effects in binary data. GEE yields Wald $\chi^2$ values as indicators of main effects and interactions. The model tested here used a binomial outcome distribution with a logit link function and a robust estimator covariance matrix.

**All Questions**

I first examined the effects of property type, domain, and age group on the likelihood of responding “yes” to the all questions (see Figure 4). Importantly, results revealed a significant main effect of age group, $\chi^2(4, N = 2,904) = 16.10, p = .003$. (Here and in subsequent reporting of GEE analyses, this $N$ refers to the total number of cases)

![Figure 4. Study 2: Proportion of “yes” responses (± 1 SE) to the all question by age group, property type, and domain.](image-url)
in the analysis.) Four- and 5-year-olds responded “yes” to the all question at significantly higher rates (M = 49.3 and 51.5%, respectively) than 6-year-olds (M = 32.1%), 7- to 8-year-olds (M = 34.1%), and adults (M = 29.5%; all ps < .05; sequential Bonferroni-corrected pairwise comparisons). One-sample t-tests comparing the number of “yes” responses against chance (50%) confirmed that 4- and 5-year-olds provided “yes” responses to the all question at chance levels (ps > .81), whereas 6-year-olds, 7- to 8-year-olds, and adults each responded “yes” less often than would be predicted by chance (ps < .01). The fact that 4- and 5-year-olds affirmed the all questions at higher levels than older participants is consistent with the prediction that young children are biased to assume within-category homogeneity in early childhood. However, this conclusion is complicated by the 4- and 5-year-olds’ chance-level performance. To further probe the nature of children’s responses to the all question, I return to the GEE analyses exploring whether participants differentiated their responses to the all question depending on property type and category domain.

In addition to the main effect of age group, GEE analyses also revealed main effects of property type, \( \chi^2(3, N = 2,904) = 112.75, p < .001 \), and domain, \( \chi^2(2, N = 2,904) = 6.72, p = .035 \), as well as interactions of Property Type x Domain, \( \chi^2(6, N = 2,904) = 61.46, p < .001 \), Property Type x Age Group, \( \chi^2(12, N = 2,904) = 45.70, p < .001 \), and Property Type x Domain x Age Group, \( \chi^2(24, N = 2,904) = 57.81, p < .001 \). To sort out these interactions, I used GEE to examine the effects of property type and domain separately for each age group. Results showed that participants at all ages differentiated their all responses by property type (ps < .001). Contrary to Study 1, even 4-year-olds, \( \chi^2(3, N = 576) = 19.60, p < .001 \), and 5-year-olds, \( \chi^2(3, N = 600) = 27.41, p < .001 \), showed significant effects of property type on their all responses. As can be seen in Figure 4, participants responded “yes” to the all questions most often in the case of category member parts, followed by behavior or function, color, or temporary or accidental properties. Comparisons of the likelihood of responding “yes” to all questions about these different property types (with sequential Bonferroni corrections) were significant in all cases for adults, 7- to 8-year-olds, and 5-year-olds (ps < .027), and in all cases except the contrast between parts and behavior or function for 6-year-olds (p = .18). Four-year-olds only showed significant differences in their all responses for parts as compared to those for other property types (ps < .01); they did not further differentiate their responses to questions about behavior or function, color, or temporary or accidental properties (ps > .74).

Finally, in considering the role of domain in responses to the all questions, analyses revealed an interaction between property type and domain in all age groups (ps < .046) except 4-year-olds (p = .16). The most consistent finding was a tendency to differentiate responses for social categories from those for animals and artifacts. Children as young as 6 were more likely to respond “yes” to all questions about the parts of social category members relative to animals (ps < .05). Likewise, children as young as 5 provided fewer “yes” responses to the all questions about the behavior of social categories relative to animals (ps < .05) and, to a lesser extent, artifacts (6-year-olds: p = .28; all other ps < .05). No domain effects were evident for 4-year-olds.

Overall, data indicate that even 4- and 5-year-olds demonstrate some ability to limit their category generalizations. Nevertheless, results also provide support for the hypothesis that a bias to assume homogeneity within categories is present early and become less pronounced with age.

**Lots and just a Few Questions**

In the next set of analyses, I examined property type, domain, and age group differences in participants’ responses to the lots and just a few questions. Importantly, whether children received these questions for a given item was dependent on their response to the prior questions: Children only received the lots question if they responded “no” to the all question and only received the just a few question if they responded “no” to the all and lots questions. Although not all children technically answered any given lots or just a few question, because endorsement of an earlier question can be considered an implied “no” response to a subsequent question, I included the full set of participants in these analyses and treated children who did not receive a given lots or just a few question as equivalent to those who responded “no” to it. Separate GEE analyses were conducted for each question (lots, just a few).

GEE analyses examining the likelihood of responding “yes” to the lots questions revealed a main effect of property type, \( \chi^2(3, N = 2,904) = 25.38, p < .001 \), an interaction of property type and domain, \( \chi^2(6, N = 2,904) = 20.04, p = .003 \), and an interaction of property type, domain, and age group, \( \chi^2(24, N = 2,904) = 42.78, p = .001 \).
$p = .011$. Data showed that, as in the case of the all questions, participants differentiated their responses to the lots questions by property type and domain increasingly with age (see Figure 5). Adults showed an effect of property type, $\chi^2(3, N = 600) = 11.64$, $p = .009$, and an interaction of Property Type $\times$ Domain, $\chi^2(6, N = 600) = 20.36$, $p = .002$ (Figure 5a). Seven- to 8-year-olds showed a similar pattern—property type: $\chi^2(3, N = 552) = 8.52$, $p = .036$; Property Type $\times$ Domain: $\chi^2(6, N = 552) = 12.31$, $p = .055$ (Figure 5b). These effects were less pronounced in 6-year-olds—property type: $\chi^2(3, N = 576) = 2.63$, $p = .45$; Property Type $\times$ Domain: $\chi^2(6, N = 576) = 12.21$, $p = .057$ (Figure 5c). Finally, 4- and 5-year-olds showed little differentiation—revealing only subtle effects of property type, 4-year-olds: $\chi^2(3, N = 576) = 7.56$, $p = .056$; 5-year-olds: $\chi^2(3, N = 600) = 10.40$, $p = .015$ (Figures 5d and 5e).

GEE analyses examining the likelihood of responding “yes” to the just a few questions revealed main effects of property type, $\chi^2(3, N = 2,904) = 159.05$, $p < .001$, and domain, $\chi^2(2, N = 2,904) = 9.00$, $p = .011$, and interactions of Property Type $\times$ Domain, $\chi^2(6, N = 2,904) = 50.90$, $p < .001$, and Property Type $\times$ Age Group, $\chi^2(12, N = 2,904) = 27.62$, $p = .006$ (see Figure 5). The three-way interaction was nonsignificant ($p = .37$). As expected, participants differentiated their just a few responses by property type, endorsing this answer most often in the case of temporary or accidental properties ($M = 55.3\%$). The effect of property type was significant for all age groups (all $ps < .016$); however, as suggested by the Property Type $\times$ Age Group interaction, this effect became more pronounced with age (see Figure 5). The most notable age effects occurred in the case of temporary or accidental properties, for which rates of agreement with the just a few question increased substantially with age, $\chi^2(4, N = 726) = 25.60$, $p < .001$. Just a few responses also varied by domain, occurring more frequently overall for social categories ($M = 30.0\%$) than for animals ($M = 23.7\%$) or artifacts ($M = 23.3\%$; both $ps = .29$). The effect of domain was most prominent for questions about behavior or function, where just a few responses were given more frequently for social categories ($M = 32.2\%$) than for animals ($M = 15.0\%$) or artifacts ($M = 13.7\%$; $ps < .001$). But, as can be seen in Figure 5, only adults and 7- to 8-year-olds were sensitive to this domain difference ($ps < .016$). Thus, results again show age-related increases in differentiation by property type and domain.

**Discussion**

Overall, Study 2 results broadly replicated the findings from Study 1 but revealed greater sophistication in the inductive inferences of young children. Specifically, data showed age-related increases in children’s sensitivity to the ways in which different kinds of properties and different kinds of categories limit property generalizability. Importantly, findings derived from this simplified task showed somewhat earlier recognition of within-category variability than the more complicated version used in Study 1. Although preschoolers still made less nuanced category inferences than older children and adults, in the current study even 5-year-olds showed some sensitivity to differences in within-category similarity based on property type and, to a lesser extent, domain. Study 2 results also provide further support for the hypothesis that young children possess a modest domain-general bias to assume similarity within categories. This bias was observed in 4- and 5-year-olds’ tendency to frequently respond affirmatively to all-quantified questions about the distribution of properties in novel category members. As in Study 1, this homogeneity bias became less pronounced with age and, in the current study, was already largely diminished by the age of 6.

Despite the modifications made in Study 2 to simplify the task demands, it is important to recognize that the inferences required in this task were still quite complex. To be successful, children needed to consider a hypothetical planet with novel categories, reflect on the likely variability among members of those categories, and then recruit and apply their conceptual knowledge about domains and property types. Study 2 also required children to interpret quantified statements about entire categories (e.g., all flooms)—a task that has been shown to be computationally challenging for preschoolers (Gelman, Leslie, Was, & Koch, 2015; Hollander, Gelman, & Star, 2002). The fact that even 5-year-olds used property type and domain to shape some of their inferences, suggests that by 5 years of age, children are capable of success on this challenging task. However, it remains possible that the homogeneity bias demonstrated by the 4- and 5-year-olds may be somewhat inflated by (a) a strategy to respond “yes” to the first question in the face of confusion or (b) a tendency to interpret the quantified question (e.g., “Are all flooms orange?”) as a generic question with considerably looser truth condition (e.g., “Are flooms orange?”; Brandone, Hedglen, & Gelman, 2015). Further research is needed.
Figure 5. Study 2: Proportion of “yes” responses (± 1 SE) to the lots and just a few questions by age group, property type, and domain.
General Discussion

The ability to successfully use category information as a basis for induction requires awareness of both the similarities and the variation that exists among members of the same category. The goal of the present research was to extend the existing data on children’s category-based inferences by examining how children use categories as a foundation for induction and appropriately constrain category inferences based on their conceptual knowledge. To do so, the current studies moved beyond classic induction measures that indicate how children generalize a property from one category member to the next and used a novel induction task assessing children’s beliefs about how properties are distributed among an entire category.

Across two studies, data suggest that children begin to differentiate their inferences based on conceptual knowledge about the nature of the to-be-generalized property and the category domain by 6–8 years of age. In particular, children’s use of abstract knowledge about the distribution of different kinds of properties was detectable by age 5 and became increasingly clear over time. Like adults, children rated the parts of exemplar category members as most likely to be shared with 100% of the category, followed by behaviors and functions, color, and temporary or accidental properties. Although prior work using traditional induction tasks has demonstrated young children’s reluctance to extend idiosyncratic or transient properties (e.g., Gelman, 1988; Graham et al., 2003), the current data are among the first to document children’s beliefs about other properties and the extent to which they are shared within categories.

Results also revealed children’s sensitivity to how the ontological domain of a category constrains induction. Although only adults showed a main effect of domain—generalizing more broadly for animals than social categories—effects of domain appeared in interaction with effects of property type by age 6 and increased over time. This evidence of children’s use of domain to limit generalizations is broadly consistent with prior work showing that preschoolers produce more category-wide generalizations about novel animal categories than novel artifact categories (Brandone & Gelman, 2009, 2013) and that by second grade children draw more inferences within natural kind categories than within artifact categories (Gelman, 1988; Gelman & O’Reilly, 1988). However, the current data also provide a more nuanced view. Children do not simply assume that all properties of naturally existing categories are more generalizable than properties of artifact categories: Domain differences also depend on the nature of the property.

Intriguingly, the clearest domain differences in the current studies emerged in the case of social categories as compared to animal categories. Social categories are unique in that they share features with both naturally occurring and constructed categories (Haslam et al., 2000; Hirschfeld, 1998; Rhodes & Gelman, 2009). For example, some social categories (e.g., gender) are viewed as natural and essentialized and thus similar in structure to animal categories; others (e.g., Midwesterners, movie buffs, and sometimes race) can be construed as conventionalized and ad hoc, and thus more similar to artifact categories (Haslam et al., 2000; Rhodes & Gelman, 2009). The induction findings reported here suggest that participants construed the target social and animal categories quite differently. For example, only the parts of social category members were generalized broadly to the entire category; other properties—including behavior and color—were not. These findings also raise questions about how children and adults interpreted the novel social categories used here. Participants were given minimal information—a single exemplar that stood as a representative of a “kind of person” on the alien planet. Thus, it was left ambiguous as to what the category referred, the level of the category (e.g., basic vs. superordinate), and whether the category was naturally occurring or more arbitrary. The fact that older children and adults were generally reluctant to generalize information beyond basic body structure (e.g., the presence of arms and eyes) suggests that they did not construe these social categories as natural and essentialized. This may reflect a general expectation about social categories—namely that they are viewed as conventionalized or ad hoc unless given information to the contrary. This may also suggest that children interpreted the social categories used in the current studies at the superordinate level—thus generalizing each property from the individual exemplar to the larger category “people” (Waxman, 2010). Further research is needed to test these possibilities. Finally, results from the current studies also showed that the social domain is the first area in which children begin to demonstrate differences in homogeneity estimates, raising the possibility that children are learning
about variability first in the case of everyday social categories and later generalizing it to other domains (see Carey, 1985 for related arguments).

In addition to documenting a refinement in beliefs about homogeneity with age, the current data also showed that early in development young children demonstrate a subtle but reliable tendency to believe that categories are homogeneous. This idea that categories mark groups of individuals that are highly similar to one another is an essential foundation of category knowledge that enables category-based induction (Gelman, 2003; Murphy, 2002). However, the bias revealed here in young children extends beyond the basic assumption present in adults. Specifically, in the current research, children below 6 years of age regularly predicted that properties of a specific novel category member were present universally in the larger category—in 100% or “all” category members. This broad view of homogeneity appeared to some extent across domains and sometimes even in unlikely cases, such as when the information being generalized involved a temporary or accidental property like being dirty or having a broken leg. Although the current data do not allow strong conclusions about this homogeneity bias because it remains possible that the complexity of the task inflated children’s 100% or “all” responses, results are nonetheless consistent with the view that young children’s category representations may be more homogeneous than those held by older children and that early in development children may operate on an assumption that categories are uniform and coherent.

Further research is needed to more accurately estimate the prevalence and breadth of 4- and 5-year-olds’ homogeneity expectations and their influence on conceptual development. However, if such a homogeneity bias exists, it could be considered adaptive in early childhood when the need to generalize information based on minimal experience is especially important. In many cases, an assumption of category homogeneity is appropriate, and thus this expectation could help to propel knowledge acquisition and explain how children readily acquire generic information about categories (Gelman, 2003; Prasada, 2000). Nevertheless, this assumption could also have negative consequences in that it may discourage children from considering and attending to meaningful variation within categories. A failure to attend to variability within categories could help to explain why children below the age of 9 typically view diverse and nondiverse samples (e.g., a Dalmatian and a poodle vs. two Dalmatians) as equally inductively powerful (see Rhodes & Brickman, 2010 for a similar argument). Moreover, assuming universal similarity within categories may also lead to mistaken inferences (e.g., concluding that all dogs are vicious after an encounter with a particularly nasty dog) that can have real consequences—especially in domains in which individuating information is highly valuable, as in the case of social categories (Berndt & Heller, 1986; Macrae & Bodenhausen, 2000; Taylor, 1996).

Despite the possible homogeneity bias observed here, young children are not completely ignorant of variability within categories. Hollander et al. (2002) have demonstrated that even 4-year-olds reject universal extensions of properties that are true in some but not all members of familiar basic-level categories (e.g., “Do all bears have white fur?”; “Do all girls have curly hair?”). Thus, young children recognize that everyday categories are not purely homogeneous. This raises the question of how the present data, which all involved novel, fictional categories, should inform our understanding of how children represent and reason about actual categories in the world about which they have some background knowledge. Data offer at least two suggestions. First, any general expectations observed here about the effects of property type and domain on induction must be abstracted from participants’ experience with real-world categories. Thus, nuanced differentiation in induction as a result of these factors implies that children are aware of how categories in the real world differ in these areas (e.g., that basic-level categories are generally homogeneous with respect to parts but less so with respect to color or that categories of animals are more likely to share behaviors than categories of people). Second, for younger children, the potential homogeneity bias demonstrated here—in spite of knowledge about variability within real-world categories—may suggest that, early in development, children approach new categories with an abstract assumption about homogeneity and that only with experience can they identify the ways in which category members differ.

A key question raised by these findings is how children come to recognize and attend to the important variability that exists within categories. If, as these data suggest, children come to category-based induction with a bias to assume that categories are essentially homogeneous, how do they acquire knowledge about the constraints on induction? One explanation for children’s growing recognition of within-category variability is a basic computational one. Young children may have difficulty representing how a single category can include multiple
varieties of a given property (e.g., having a handle or not; eating leaves or bugs). As children's computational resources increase with age, they may become better able to handle variability within categories. A second (not mutually exclusive) possibility highlights the role of experience in shifting children's expectations about variability within categories. When young children are new to category learning, they may have fewer experiences with distinct or atypical category instances. Over time, children may be faced with greater counterevidence as they acquire more experience with dissimilar category members. Such opportunities may lead to shifts in representations of familiar categories that become abstracted over time and thus generalizable to novel categories. Additional research is needed to test these predictions and to shed light on the mechanisms underlying the age-related increases in awareness of variability within categories.

The novel empirical approach to category-based induction developed in the current studies also has some important implications. Unlike prior methods in which children were asked to generalize a property from one category member to another, in the current studies, children were asked to use a single category member to predict the distribution of properties among the entire category. This approach provides unique insight into the issues of within-category homogeneity and variability. First, results go beyond the conclusion that young children generalize information more broadly than older children, and demonstrate that in some cases preschoolers specifically expect novel categories to show 100% homogeneity. Second, these data suggest that around age 7, children further distinguish between high-prevalence scenarios (showing distinct expectations about 100% and 80% or "lots" distributions) and between low-prevalence scenarios (showing distinct expectations about 50% and 20% or "few" distributions). Finally, by allowing for more nuanced prevalence estimates, these tasks reveal subtle differences in children's expectations and knowledge by property type and domain. Although there are limits to this approach—most clearly its complexity and computational demands—the current findings offer useful, novel contributions to our understanding of children's category representations.

A final question to consider is how the current findings speak to the lively theoretical debate in the literature regarding whether young children use a knowledge-based (Gelman, 2003; Gelman & Davidson, 2013; Gelman & Waxman, 2007) or similarity-based (Badger & Shapiro, 2012; Sloutsky & Fisher, 2004; Sloutsky et al., 2007) approach during category-based induction. The present studies did not pit perceptual appearance against category membership as is typical in studies in this literature (e.g., Gelman & Markman, 1986); instead, the current approach was to begin with the premise that a category exists and then to assess the extent to which children expect additional category members to share both perceptual (e.g., parts, color) and non-perceptual (e.g., behavior or function) features. Given these methodological differences, results cannot speak directly to this debate. Nevertheless, connections can be drawn. First, evidence of young children's homogeneity expectations can be seen as consistent with the knowledge-based account, which proposes that, from early in development, induction is driven by conceptual knowledge and the belief that individuals belong to categories that share important properties (e.g., Gelman, 2003). If a homogeneity bias in fact exists, it can be considered strong evidence for this knowledge-based account. However, some aspects of the current results are also consistent with the similarity-based account, which argues that category knowledge is a product rather than a precondition of development and that conceptual knowledge develops beyond early childhood as a result of a protracted process of perceptually driven domain-general learning (e.g., Sloutsky & Fisher, 2004). Specifically, support for this view comes from the current evidence of clear developmental differences in induction, including increases in children's use of property type and category domain information beginning around age 6. Overall, the current studies can be seen as further evidence for a need to move beyond this debate toward a more nuanced view of category-based induction that recognizes the role of both early expectations and developing knowledge (see also Booth, 2014; Gelman & Davidson, 2013).

In conclusion, the vast amount of knowledge humans possess about the world depends on our ability to successfully generalize category information beyond our own limited experience. The current research sheds light on one foundational component of this process—the understanding that categories encompass a subtle balance between shared characteristics that are common across category members and distinct characteristics of individuals. Results suggest that knowledge about when and how to limit inferences during category-based induction emerges early and becomes increasingly sophisticated across childhood. Data also provide tentative support for the view that young children come to the task of category-based
induction with an early developing assumption that categories are largely homogeneous. These findings raise important questions regarding the mechanisms underlying children’s growing recognition of variability within categories and the consequences of the emerging balance between understanding similarity and variability for children’s conceptual development.

References


Appendix

Full Set of Items and Properties by Domain and Property Type

<table>
<thead>
<tr>
<th>Property type</th>
<th>Animal</th>
<th>Artifact</th>
<th>Social category</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parts</td>
<td>Has a tail</td>
<td>Has a cord</td>
<td>Has arms</td>
</tr>
<tr>
<td></td>
<td>Has ears</td>
<td>Has a handle</td>
<td>Has eyes</td>
</tr>
<tr>
<td>Behavior/function</td>
<td>Slides on its belly</td>
<td>Gets pushed around</td>
<td>Walks backwards</td>
</tr>
<tr>
<td></td>
<td>Eats ladybugs</td>
<td>Is used to reach things from high places</td>
<td>Eats flowers</td>
</tr>
<tr>
<td>Color</td>
<td>Is blue</td>
<td>Is red</td>
<td>Has pink hair</td>
</tr>
<tr>
<td></td>
<td>Is orange</td>
<td>Is blue</td>
<td>Has green skin</td>
</tr>
<tr>
<td>Temporary/accidental</td>
<td>Is dirty</td>
<td>Is dirty</td>
<td>Is dirty</td>
</tr>
<tr>
<td></td>
<td>Has a broken tail</td>
<td>Has a broken handle</td>
<td>Has a broken leg</td>
</tr>
</tbody>
</table>

Social category images adapted from iStock.com/Bubert.