NIH Public Access
Author Manuscript
Cogn Psychol. Author manuscript; available in PMC 2011 November 1.

Published in final edited form as:
Cogn Psychol. 2010 November ; 61(3): 273–301. doi:10.1016/j.cogpsych.2010.06.001.

Effects of Generic Language on Category Content and Structure

Susan A. Gelman,
University of Michigan

Elizabeth A. Ware, and
Viterbo University

Felicia Kleinberg
University of Michigan

Abstract

We hypothesized that generic noun phrases ("Bears climb trees") would provide important input to children's developing concepts. In three experiments, four-year-olds and adults learned a series of facts about a novel animal category, in one of three wording conditions: generic (e.g., "Zarpies hate ice cream"), specific– label (e.g., "This zarpie hates ice cream"), or no-label (e.g., "This hates ice cream"). Participants completed a battery of tasks assessing the extent to which they linked the category to the properties expressed, and the extent to which they treated the category as constituting an essentialized kind. As predicted, for adults, generics training resulted in tighter category-property links and more category essentialism than both the specific-label and no-label training. Children also showed effects of generic wording, though the effects were weaker and required more extensive input. We discuss the implications for language-thought relations, and for the acquisition of essentialized categories.

Much of children's knowledge about the world is obtained not through direct experience but through the testimony of others (Gelman, 2009; Gergely & Csibra, 2006; Harris & Koenig, 2006). For example, many scientific concepts (shape of the earth, evolution), religious concepts (qualities of the supreme being, reincarnation beliefs), social concepts (caste, nationality), and conventions (appropriate school attire, meaning of traffic lights) cannot be discovered by a child wholly on his or her own, but require interacting with others. It has long been observed that language is a primary means of transmitting information from one generation to the next, and thus of guiding thought (Bowerman & Levinson, 2001; Gentner & Goldin-Meadow, 2003; Vygotsky, 1934/1962). In the present paper, we consider the role of one particular aspect of language, that of generic noun phrases.

Generic noun phrases (e.g., "Bears climb trees"; see Gelman, 2004) potentially provide an important source of information to children's developing concepts, in that they refer directly to *kinds* of things. Consider the distinction between "this bear climbs trees" and "bears climb trees". The former (non-generic) refers to a present, perceptible object (a single bear), whereas the latter (generic) refers to an abstract category (bears in general). One can learn all sorts of things about individuals by inspecting them on one's own, for example, by observing a single bear climbing a tree; one cannot so easily learn about kinds without the input and guidance of others. For example, one cannot directly observe 'bears,' as a category, climbing trees.

In particular, there are two sorts of facts that generics may convey. First, generics imply a relatively close link between a category and a property. For example, "Bears climb trees" implies that bears typically or in general climb trees (Gelman, Star, & Flukes, 2002). In contrast, "This bear climbs trees" tells us that one particular bear has the property of climbing trees, but leaves open how broadly to extend this property – it could be true of just this particular bear, or true of bears more generally. Likewise, viewing a bear climbing a tree without any linguistic context leaves open the possibility that this property could be either idiosyncratic of this bear or true more generally of bears. Generics furthermore often imply that particular features are not only frequent but also relatively central to a category (e.g., that climbing trees isn't an accidental feature of bears, but a normative expectation; Cimpian & Markman, in press; Prasada & Dillingham, 2006, 2009).

Second, generics may also suggest that a given category is richly structured (e.g., that bears share stable commonalities, that there are strict boundaries between bears and other animals, that bears have innate qualities). That is, hearing generics expressed about a category may convey an essentialist view of that category. Psychological essentialism is the intuitive belief that certain categories have an underlying reality that cannot be observed directly but that gives an object its identity, and is responsible for other similarities that category members share (Gelman, 2003; Medin, 1989). For biological concepts, an essence is whatever quality remains unchanging as an organism grows, reproduces, and undergoes morphological transformations (e.g., from baby to adult). Psychological essentialism has two distinct components: a natural kind component (that certain categories have a cluster of non-obvious, inherent properties) and an essence component (that an internal part or quality causes the commonalities shared among members of the kind) (Gelman, 2003). In the current research, we focused exclusively on investigating the first component of psychological essentialism, that certain categories have a wealth of non-obvious, inherent properties. We did not specifically examine the further question of whether people attribute an inner, causal part or substance to all category members (see Strevens, 2000; Ahn, Kalish, Gelman, Medin, et al., 2001, for debate). Therefore, our use of the term "essentialism" henceforth refers to the first component.

Although it has been argued that people have a tendency to view categories in essentialist terms, it is also clear that categories vary in this respect. For example, gender is readily essentialized from childhood onward but race is inconsistently essentialized, depending on a person's age and cultural context (Kinzler, Shutts, DeJesus, & Spelke, in press; Rhodes & Gelman, in press; Waxman, in press). Generic language input may be one source of information that guides children's construals.

There are several reasons to suppose that children as young as preschool age might be sensitive to these implications of generics for categories. First, generics are frequent in child-directed speech (Gelman, Coley, Rosengren, Hartman, & Pappas, 1998), and are understood appropriately by preschoolers (Gelman & Raman, 2003; Hollander, Gelman, & Star, 2002), thereby suggesting that generics have the potential to influence children's conceptual representations. Second, generics seem to be linked to important aspects of conceptual structure. For example, children and adults alike produced generics more often to describe animal categories (such as duck or horse) than artifact categories (such as chair or spoon; Brandone & Gelman, 2009; Gelman, Goetz, Sarnecka, & Flukes, 2008; Goldin-Meadow, Gelman, & Mylander, 2005). Likewise, 3- and 4-year-olds children are more likely to interpret an ambiguous sentence as generic (e.g., "They are afraid of mice" in the context of two birds), if they have prior knowledge that the given property being is central to the category (Cimpian & Markman, 2008). Finally, children as young as 3 years of age correctly recall whether sentences are provided in generic or non-generic form (Gelman & Raman, 2007), demonstrating that generic input is stored in long-term memory. These component pieces all suggest that generics are semantically important, acquired early, and have the potential to influence children's conceptual representations.

In support of this hypothesis, there is initial evidence demonstrating that generic language actually influences children's concepts (versus merely reflecting their concepts). Specifically, generics strengthen the links between a category and properties attributed to it, for children as young as 4 years of age (see Chambers, Graham, & Turner, 2008; Cimpian & Markman, in press; Hollander, Gelman, & Raman, 2009). For example, learning that "blickets have stripes" increases children's tendency to infer that a new striped animal will be a blicket, as well as their tendency to infer that a new blicket will have stripes. Furthermore, generics also increase children's focus on deeper conceptual relations, so that being asked to compare "dogs" and "cats" leads to a focus on less obvious features in contrast to being asked to compare a particular dog with a particular cat (Gelman, Raman, & Gentner, in press).

The Present Research

The present research further examines whether and how generic language influences children's concepts. It extends beyond prior work in two primary respects. First, whereas prior work provided minimal generic input (e.g., typically just one property per category) and tested immediate effects (e.g., extension of the property to novel instances), the current research provides extensive generic input (17 properties per category, in some cases repeated numerous times) and tests effects following a delay (from several minutes to two weeks). We seek to discover whether generics influence children's concepts in an enduring way. The repeated input is consistent with how novel categories are often introduced in an informational book or TV program (e.g., Gelman et al., 1998). The introduction of a delay allows us to assess whether generic input has effects beyond its immediate semantic implications.

A second way that the present work extends prior work concerns the kinds of effects that are examined. As noted earlier, prior work has tended to focus on the strength of the link between the target category and the properties expressed generically. For example, a child might learn "blickets have stripes" in the context of one blicket, and then be asked whether other blickets also have stripes, and whether another animal with stripes might also be a blicket. A largely untested additional question is whether generics foster a more essentialized view of the target category. For example, if a child hears "blickets have stripes", does this increase her inference that blickets are a stable category, with non-obvious similarities, innate features, that promotes inferences beyond the stated properties? The present studies examine both of these hypothesized effects: category-property links and essentialism.

A further purpose of the research is to examine the role of generics relative to the role of labeling in concept formation. On the one hand, common noun labels (e.g., "This is a bird") express membership in a kind. Thus, labeling, in and of itself, encourages a focus on categorization (Fulkerson & Waxman, 2007; Waxman & Markow, 1995) and may encourage children to treat a category as a richly structured kind (Cimpian, Arce, Markman, & Dweck, 2007; Gelman & Heyman, 1999; Xu, 2005; Yamauchi, 2005). Moreover, labels can express category membership even when they compete with misleading perceptual features (e.g., an atypical mammal, such as a bat, is still a mammal). Labels therefore imply a distinction between deeper properties (what something *is*) and surface properties (what something *is like;* Wierzbicka, 1994). On the other hand, beyond the effects of labeling per se, generics might imply greater stability and essentialism. The present studies tested this prediction, thus permitting a test of competing hypotheses regarding the role of language in concept formation.

To investigate these issues, we conducted a series of three studies that examined the effects of providing generic input regarding a novel animal category. Study 1 tested college students who carefully read through a picture book that provided extensive input about a single category.

The wording of the informational text in the book varied systematically across three conditions: *generic*, *specific-label*, and *no-label*. Each page of the book described a distinct property of either an individual exemplar (specific-label and no-label) or the category as a whole (generic). Participants next completed a battery of tests that measured category-property links and essentialist beliefs about the category. Testing occurred after a several-minute delay to ensure that any effects of condition did not stem from immediate (short-term) recall. In Study 2, we tested preschool children and their parents, using the same picture book and wording conditions. However, in order to boost recall of the information in the book, children heard the book repeatedly over the course of a week, when parents read the book to the child daily at home. Testing occurred at two time points: several minutes after the final book reading (which occurred at the end of the week during which participants repeatedly read the book), and after a two-week delay following the final book reading. Study 3, with preschool children and their parents, differed from the first two studies in providing minimal and piecemeal input and no delay. Rather than receiving repeated input all at once, participants received a single input sentence for each task, followed by immediate testing based on that one sentence.

We predict that generics will affect children's and adults' representations of the novel concept, leading to two sorts of effects: (a) greater links between the category and the properties that are expressed, and (b) more essentializing of the novel category (in the sense of treating the novel category as a natural kind with inherent, non-obvious properties). We predict these effects will obtain in comparison to the specific-label condition as well as in comparison to the no-label condition. Furthermore, we predict that the specific-label condition will lead to greater essentialism than the no-label condition. Thus, we predict step-wise effects, whereby providing a label will lead to greater category coherence than when no label is provided, but providing generic input will be even more powerful than a label alone.

Within this basic framework, three additional questions are key. First, we are particularly interested in the effects of participant age, and specifically whether young children are sensitive to the implications of generic wording, despite their more limited experience with the distinctions between different linguistic forms. A second key question is whether any effects of generics and labeling obtained after a brief delay will also persist after a more extended delay. This question is addressed in Study 2. Finally, a third question is whether cumulative effects are greater than piecemeal effects. That is, does providing a series of input sentences in a row (as in a picture book) lead to different effects than hearing just one sentence at a time? This question is addressed by comparing the results of Study 3 to those of Studies 1 and 2.

Study 1

Methods

Participants—42 adults (30 females, 12 males) participated. Most were undergraduates at a large, Midwestern university (*M* age = 20.5 years). They received financial compensation for their participation. Participants were randomly assigned to condition (14 each in the generic, specific-label, and no-label conditions).

Materials and Procedure—Participants were tested individually in an on-campus lab. First, they were given five minutes to read a picture book, which provided training (henceforth referred to as the "training book"). They were instructed to read the words on each page and to use the entire five minutes. This was done to ensure that participants paid careful attention to the wording in the book. Then the book was removed and there was a four-minute delay. During the delay participants completed a distracter task (to prevent rehearsal of the book material), in which they were asked to solve multi-digit multiplication problems. Then they completed a battery of test items (see Appendix), which they were asked to fill out at their own pace. Finally, they completed a 140-item questionnaire that measured their general endorsement of essentialist beliefs. The questionnaire asked participants to rate (on a 5-point Likert scale) stability, innateness, inheritance, genetic causes, detection in infancy, environmental influences, and capacity to change for each of 20 human traits (e.g., shy, musically talented, neat, intelligent) (see Gelman, Heyman, & Legare, 2007;Haslam, Rothschild, & Ernst, 2000).

The training book focused on a single category, and each page provided a different category exemplar exhibiting a different property. Drawings and their corresponding text were on facing pages. The novel category was “zarpies”, which are fictitious four-legged animals (presumably mammals), large in size, with horns and brightly colored fur (see Figure 1 for examples). We chose to focus on an animal category, as generics are most commonly used in this domain (Brandone & Gelman, 2009).

The book included 17 pictures (plus cover), with accompanying text. Table 1 lists the properties, in the order they appeared in the book. The properties were chosen as ones that could potentially be either idiosyncratic to that individual or generalizable across the kind. This was important so that the wording manipulation could possibly have an effect (i.e., to avoid floor or ceiling effects). We also designed the properties with the test battery in mind. The properties had to meet certain criteria needed for testing. (For example, in order to test whether participants would categorize a new instance based on a property expressed for one exemplar in the book, we needed some properties that could be displayed but would not be visible in most of the pictures, such as stripes on the bottom of feet. In order to test innateness beliefs, we needed some properties that contrasted with the adoptive-parent category of “dogs”; e.g., like to sing vs. like to bark.) In order to vary the properties, roughly equal numbers of psychological, behavioral, and physical features were included. Finally, the properties were designed to be child-friendly, amusing, distinctive, and memorable.

There were 20 zarpies depicted in the book: 3 on the cover and 1 in each of the following 17 pictures. The different instances all had the same overall body shape and the same number and kind of parts. The exemplars varied systematically along five dimensions: fur type, leg/foot size, horn shape, tail type, and snout shape. Each dimension had one prototypical value (long fur, short legs/wide feet, curved horns, pig-like tail, average snout) displayed by 60% of the instances in the training book, as well as two atypical values (each displayed by 20% of the instances in the training book). Thus, the items can be considered to consist of variations around a prototype (see Ward, Vela, & Hass, 1990, for similar item structure), although the prototype itself was never shown to participants. Each category exemplar matched the prototype on 3 of these 5 features (e.g., one exemplar had long fur, curved horns, and a pig-like tail). There were also two features (fur color, fur pattern) that varied randomly: each of these features had three values, which were roughly evenly split across the instances in the training book. Table 2 details the structure of the items in the book; each number represents a value on the relevant dimension (e.g., fur type: long, short, or curly). The value of “1” indicates the prototype for all dimensions except fur color and fur pattern (which, as noted earlier, vary randomly and do not have a prototypical value).

The illustrations and properties were identical across the 3 conditions, differing only in the wording (see Table 3 for summary). Each page of text began with an orienting statement referring to the instance depicted on the facing page, either “Look at this zarpie!” (generic and specific-label conditions) or “Look at this!” (no-label condition). Then a target property was provided. In the generic condition, a bare plural, generic noun phrase was used (e.g., “Zarpies like to sing”). In the specific-label condition, a definite singular noun phrase was used (e.g., “This zarpie likes to sing”). And in the no-label condition, the animal was referred to using only the demonstrative pronoun (e.g., “This likes to sing”). The book title (written across the cover) also varied across conditions: *All about Zarpies!* (generic), *Look at This Zarpie!* (specific-label), or *Look at This!* (no label).

The tasks in the test battery (excluding the memory questions) are presented in the Appendix. Participants completed a paper-and-pencil version of the test battery. The core questions were designed to assess two aspects of zarpie concepts: (a) the strength of the link between the category and the properties expressed in the training book, and (b) the degree to which the category is essentialized (e.g., treated as stable, inductively rich, having innate qualities). We included 4–6 items per task to balance two competing factors: the tasks are extensive enough to obtain a stable estimate of performance, but brief enough to ensure that the entire testing session was not overly long, roughly 20 minutes in length. Additional questions were included to assess (c) participants' memory of the training items.

As noted above, the wording of the training book differed systematically across conditions during training. In contrast, the *test* questions were designed to be as equivalent as possible across conditions. All test items referred to specific instances (e.g., "Why is this chasing a shadow?" or "Can you find another zarpie/one?"). The wording of the test questions was identical in the generic and specific – label conditions, typically referring to "this zarpie" or "another zarpie". The test questions in the no-label condition were the same except that they never mentioned the label "zarpie", instead, typically referring to "this" or "this one" or "another one".

A brief description of each task in the test battery and our predictions are below, along with information about coding.

Category-Property Link Tasks

Categorization: On each trial, participants were shown a category exemplar from the training book and were asked to "Find another zarpie" (generic and specific-label) or "Find another one" (no-label), given two choices. One choice was more similar overall to the target and the prototype but lacked a critical feature mentioned on one page in the book (e.g., striped feet); the other was less similar overall to either the target or prototype, but displayed the critical feature. If participants linked the learned property to the category, then they should extend the category based on the property rather than overall perceptual similarity. Each property-based response received a "1"; each non-property-based response received a "0". Scores were summed. There were four trials. Scores could thus range from 0–4.

Familiar property induction: On each set, participants saw an instance from a book and were reminded of the property expressed on that page in the book (e.g., "This zarpie hates ice cream" (generic and specific-label); "This hates ice cream" (no-label). They were then asked whether the property generalizes to each of two new instances (presented one at a time) (e.g., "Do you think that this (zarpie) hates ice cream?"). The new instances differed from the original in color, pattern, and at least two additional features. For example, in one set, the original zarpie had long purple fur with spots, and the test zarpies had short blue, striped fur and short, curly, pale-green fur, respectively. Furthermore, both test items possessed different tail types and leg lengths from the original. This task also tests the strength of the link between the category and a property mentioned in the book. If people link the category and the property, they should extend the familiar property to novel exemplars. There were three sets of items, each consisting of an instance from the book and two new instances, amounting to six trials total. Each "yes" response received a "1"; each "no" response received a "0". Scores were summed and could range from 0–6.

Explanation (scope): On each trial, participants were provided with instances of an animal from the book and were asked why it was engaging in the target behavior (e.g., "Why is this sleeping in a tall tree?") (same wording for all three conditions.) One set of analyses for this task focused on how often participants appealed to generic scope in their explanations (e.g., "They…" or "Zarpies…"). See Table 4 for further examples of explanation coding. If people perceive a tight link between the category and the expressed properties, they should more often appeal to generic scope in their explanations (Cimpian & Markman, in press;Hollander, Gelman, & Raman, 2009;Prasada, 2000). Each generic response received a "1"; each non-generic response received a score of "0". There were four trials. Scores were summed and could range from 0–4.

Essentialism Tasks

Novel property induction: On each set, participants learned a novel property (not in the training book) of a new zarpie (e.g., "This zarpie likes to play with Frisbees" (generic and specific-label); "This likes to play with Frisbees" (no-label)), and were tested on how readily they generalize the property to other, new zarpies. This task was identical to the familiar property induction task, except that the property was not one that had been provided in the book. Thus, participants never heard this property in generic form. If people are reasoning about this category from an essentialist perspective, then they should extend even a novel property to new exemplars. Each "yes" response received a "1"; each "no" response received a "0". Scores were summed and could range from 0–6.

Stability: On each set, participants saw a familiar category instance, were reminded that a certain property is true of it at 8 years of age (e.g., "This zarpie likes to sing" (generic and specific-label); "This likes to sing" (no-label), and then were asked whether the property was true when the animal was 4 years old, and whether it would always be true (task from Gelman & Heyman, 1999). These questions test how stable a property is over time. If people are essentialist about the category, then they should expect the property to persist. Each "yes" response received a "1"; each "no" response received a "0". There were three sets and two questions per set ("4 years old" and "always"). Scores for each question were summed and could range from 0–6.

Inheritance: Participants heard nature-nurture vignettes, each about an animal whose birth-mother was a zarpie and whose adoptive-mother was a dog. They were then asked whether the animal would behave like a zarpie or a dog (e.g., eat tulips or chew bones). This task measures beliefs about innateness. If people are essentialist about the category, they should judge that the baby will inherit properties from the zarpie, regardless of upbringing. Each birth-parent response received a "1"; each adoptive-parent response received a "0". There were four vignettes. Scores were summed and could range from 0–4.

Explanation (content): As noted earlier, participants were provided with instances of an animal from the book and were asked why it was engaging in the target behavior (e.g., "Why is this sleeping in a tall tree?"). A second set of analyses of the explanations focused on how often participants appealed to an intrinsic quality or disposition of the animal ("because he likes sleeping in a tree") rather than an idiosyncratic reason ("because he's tired"). See Table 4 for further examples of explanation coding. If people are essentialist about the category, they should more often judge that the reasons for these properties are intrinsic. Each intrinsic response received a "1"; each non-intrinsic response received a "0". Scores were summed and could range from 0–4.

Taxonomic vs. thematic vs. shape: On each trial, participants were shown a familiar instance from the book, learned a new word for it, and were asked to extend it to one of three instances: taxonomic match (e.g., flying bat), shape match (e.g., racecar shaped like the target zarpie), or thematic match (e.g., fence) (see Figure 2). (This task was modeled after Imai, Gentner, &

Uchida, 1994.) If people are essentialist about the category, they should extend the new word taxonomically. Each taxonomic response received a “1”; each non-taxonomic (thematic or shape) response received a “0”. There were four trials. Scores were summed and could range from 0–4.

Memory Tasks

Picture memory: On each trial, participants were shown two similar pictures, were told that one had been in the book that they had read previously, and were asked to indicate which of the two had been in the book. Each correct response received a “1”; each incorrect response received a “0”. There were four trials. Scores were summed and could range from 0–4.

Book-reading: Participants were given a copy of the book, identical to the one that they had read previously, except that there were no words on the pages. They were asked to write down the words that had been on the pages. The purpose of this task was to check whether participants remembered the form in which the information had been provided, thereby validating the experimental manipulation. This also serves to check that participants attended equally well to the wording in the three conditions. Each correctly recalled subject noun phrase (e.g., “Zarpies”, “This zarpie”, “This”) received a score of “1”, and each incorrectly recalled subject noun phrase received a score of “0”. These scores were summed and could range from 0–17. The predicates were scored as well, with each informational piece receiving a “1” if recalled correctly. Thus, a given sentence could have several pieces (e.g., “… wiggle their horns when they are happy” had 3 pieces: wiggle / their horns / when they are happy). The predicate scores were summed and could range from 0–58. The total book-reading score (subject nouns and predicates) could range from 0–75.

Task order: The tasks were presented in a fixed order that was designed to minimize carry-over from one task to another. Specifically, participants received the tasks as follows: picture memory, explanation, stability, inheritance, categorization, familiar property induction, novel property induction, taxonomic vs. thematic vs. shape, and book reading.

Scoring and Task Intercorrelations—Three composite scores were compiled: a Category-Property composite (the sum of all scores from the categorization, familiar property induction, and explanation scope tasks; these scores could range from 0–14), an Essentialism composite (the sum of all scores from the novel induction, explanation content, stability, inheritance, and taxonomic-thematic-shape tasks; these scores could range from 0–24), and a Memory composite (the sum of all scores from the picture memory and book-reading tasks; these scores could range from 0–79). Although the composites were created on the basis of a priori theoretical reasons, we also examined the extent to which scores intercorrelated within each composite, using Pearson correlations for all analyses. For the Category-Property composite, all three tasks significantly intercorrelated, as predicted (Pearson correlations ranging from .41-.70, *p*s < .01). For the Essentialism composite, the coding of explanations (content) significantly correlated with stability (.34, $p < .03$). The lack of a single cluster of intercorrelated tasks is consistent with prior research showing that essentialism is multifaceted (Gelman, Heyman, & Legare, 2007; Haslam, Rothschild, & Ernst, 2000). For the memory composite, none of the tasks significantly correlated with one another, suggesting that the memory tasks tap into distinct constructs of memory in this context.

Results

Two coders scored the explanation (content) task responses, with 20% ($n = 9$) of the participants’ responses coded by both coders to assess reliability. Reliability was high (97%, K = .94). Two additional coders scored the book-reading task responses, with 33% ($n = 14$) of the participants' responses scored by both coders to assess reliability. Reliability was high (98%, K =.95). Means for all task composites are presented in Figure 3.

We present the results in phases. First, we examine responses to the essentialism questionnaire, to determine whether the conditions were equivalent in these background beliefs. Next, we report separate ANOVAs for the category-property composite, the essentialism composite, and the memory composite, with the primary question of interest being whether condition differences obtain.

Essentialism questionnaire—Responses to the essentialism questionnaire were scored such that 1 indicated the least essentialist endorsement and 5 indicated the most essentialist endorsement. Scores were then summed across all 140 questions. These data were then entered into a one-way ANOVA, with condition (generic, specific-label, no-label) as the between-subjects factor. There were no significant differences between conditions, $F(2,39) = 0.07$, n.s. (generic $M = 3.37$; specific-label $M = 3.42$; no-label $M = 3.41$). Because scores could range from 1–5, the scale midpoint is 3.0. Thus, participants were on average near the midpoint, but somewhat more likely to endorse essentialist responses than not. Importantly, then, participants in the different conditions did not differ in their overall essentialism beliefs, permitting us to assume that any differences across conditions on the test battery stemmed from the training book wording.

Category-property composite—This analysis examines the composite score assessing the strength of the connection between the target category (zarpies) and the properties expressed in the training book. A one-way ANOVA was conducted, with condition (generic, specific-label, no-label) as the between-subjects factor. As predicted, there was a significant effect of condition, $F(2,39) = 73.00$, $p < .001$, $\eta^2 = .79$. (All η^2 results that we report use the partial η^2 formula (SSeffect/(SSeffect + SSerror)). Tabachnick & Fidell (1989) suggest that partial η^2 is an appropriate alternate computation of η^2.) Planned *t*-tests indicate that the category-property links were stronger in the generic condition than in the specific-label and no-label conditions, *p*s < .001, which did not differ significantly from one another. When examined alone, all three individual tasks showed significant condition effects in the same direction as the composite effect just described: categorization, familiar induction, and explanation (scope), all *p*s < .001. For each of these tasks, scores in the generic wording condition were significantly higher than scores in each of the other two conditions, which did not differ significantly from one another.

Essentialism composite—This analysis examines the composite score assessing essentialist beliefs regarding the target category (zarpies). A one-way ANOVA was conducted, with condition (generic, specific-label, no-label) as the between-subjects factor. As predicted, there was a significant effect of condition, $F(2,39) = 22.09$, $p < .001$, $\eta^2 = .53$. Planned t-tests indicate that essentialism scores were significantly higher in the generic condition than in the specific-label condition and no-label conditions, and that scores in the specific-label condition were significantly higher than in the no-label condition, *p*s < .01. When examined alone, four of the five individual tasks showed condition effects: explanation (content) ($p < .05$, with generic higher than the other two conditions), stability ($p < .01$, with generic higher than no-label, and a trend showing generic higher than specific-label), novel induction ($p < .02$, with generic higher than the other two conditions), and taxonomic vs. thematic vs. shape ($p < .001$, with generic higher than specific-label which is higher than no-label). Additionally, inheritance showed a trend toward a main effect of condition ($p < .10$, with generic higher than no-label).

Memory composite—This analysis examines participants' memory for information in the training book. A one-way ANOVA was conducted, with condition (generic, specific-label, no-label) as the between-subjects factor and the memory composite score as the dependent variable. As predicted, there were no significant effects for condition, $F(2,39) = 0.07$, n.s.

Participants in the three conditions did not differ in their memory for the information in the book (pictures or text). As can be seen in Figure 3, recall was highly accurate across all three conditions, with participants recalling over 75% of the relevant information (pictures and text) provided in the books. Importantly, most errors concerned details of the predicates. Memory for the target noun phrase (which was central to the condition manipulation) was uniformly excellent (M = 92%). When examined alone, none of the individual tasks showed significant condition effects.

Discussion

The results of this study demonstrate that generic wording had at least two primary effects on adults' category learning: it strengthened the link between the category and the particular properties that were expressed in the input and, more generally, heightened an essentialist construal of the category as a natural kind. These results conform to our prediction that generics guide category formation and essentialization of categories. However, as we discuss below, the effects of wording varied somewhat for category-property links versus essentialism.

Generic wording clearly led to tighter links between the category and the properties expressed; for example, "Zarpies hate ice cream" is construed to imply that a range of zarpies (zarpies in general) have this property. Thus, when participants heard "Zarpies hate ice cream" in the context of one particular zarpie, they did not restrict this property to the instance at hand, but instead readily generalized this fact to other zarpies, even those that were completely novel (familiar property induction task). Interestingly, though, prior work has shown that generics (unlike quantifiers) need not be true of the majority of instances (Cimpian, Gelman, & Brandone, in press; Leslie, 2008). For example, we accept that "Birds lay eggs" although only female birds do so, and we accept that "Sharks attack swimmers" although few do. Thus, generics do not literally entail that the stated predicate is true of the majority of category instances (unlike a quantifier such as "most"). Nonetheless, they strongly imply for adults that the property can extend to novel category instances.

Likewise, when participants heard, "Zarpies have stripes on the bottom of their feet" in the context of one particular zarpie, they readily extended the category to new instances on the basis of this information (i.e., in the categorization task, they were more likely to classify a new instance as a zarpie if it had stripes on the bottom of its feet). Finally, when participants heard, "Zarpies sleep in tall trees", they were more likely to appeal to this generalization when asked to explain why a particular zarpie is sleeping in a tall tree (explanation task (scope)). All of these effects showed a sharp distinction between the generic condition and the other two conditions (specific-label and no-label). These results are consistent with other recent findings showing that generics tighten the link between a category and the property expressed (Chambers et al., 2008; Hollander et al., 2009). It is notable and intriguing, however, that in the current study these differences obtained even though the test questions were framed non-generically. Thus, the prior experience with the training book — not the wording at test — guided participants' responses.

Perhaps more striking, generics also influenced beliefs about the structure of the category more broadly (beyond the particular properties expressed during training). Thus, hearing generics led participants to treat the category as more stable, more inductively rich, etc. For example, hearing that zarpies hate ice cream, can bounce a ball on their backs, have stripes on the bottom of their feet, etc., leads to a general expectation that zarpies will share new properties never before encountered, that zarpies will maintain certain properties over time, and that these properties will emerge even in the face of conflicting environmental input. These effects showed differences across all three conditions: generics led to higher scores than the other two conditions, but also simply hearing a category label led to more category essentialism than hearing no label at all.

As noted above, the generic condition led to higher scores for both category-property links and essentialism. Importantly, these effects hold up compared to both of the other wording conditions: both the no-label condition (in which none of the zarpies were labeled as such) and the specific-label condition (in which the zarpies were labeled, but all properties were attributed to individuals rather than kinds). This result is important, as it demonstrates that the generic advantage is not due to labels per se, but rather to the inclusion of kind-referring expressions. However, in this study, simply providing a label (compared to the no-label baseline) did lead to greater category essentialism as well, consistent with the idea that labels do important conceptual work in implying category coherence (Waxman, 2004).

The condition effects were not due to differences in memory for the pictures or the text provided during training, given that participants performed equally well across the conditions in the memory tasks. In all three conditions, participants showed excellent recall of the pictures and of the text in the training book.

The findings from Study 1 provide a foundation for asking when these wording effects develop. Are they an outgrowth of many years of experience with generic wording, and thus obtained among adults only, or instead do they appear and guide category learning in early childhood? We addressed these questions in Study 2.

Study 2

Study 2 was designed to examine the effects of generic wording among preschool children and their parents. Additionally, the training period was more extensive (multiple readings over the course of 7–14 days), and testing was conducted twice: once immediately after the training period, and a second time 1–2 weeks later, with no intervening input after the first testing session. The more extensive training period was designed to provide the children with greater experience with the input language, in order to ensure sufficient learning and recall of the training materials with this much younger age group. The repeated testing design permits an examination of whether wording effects, if obtained, persist over time.

Methods

Participants—48 preschool children (22 females, 26 males; mean age, 55 months; age range, 49–61 months) and 48 parents (46 females, 2 males) participated. Most dyads were White ($n = 44$). Participants were randomly assigned to condition (16 dyads each in the generic, specific-label, and no-label conditions; mean ages of children were 54 months, 55 months, and 55 months, respectively). An additional six dyads participated but were excluded because they did not complete all three visits ($n = 3$), there was too great a lapse between visits 1 and 2 ($n = 1$), or there was experimenter error ($n = 2$).

Materials—Materials included those of Study 1. In order to administer an oral version of the test task battery to children, the test items were printed on laminated cards (1–3 pictures per card depending on the task). Additional materials were the Peabody Picture Vocabulary Test (PPVT) and a literacy questionnaire that assessed the frequency of reading activities in their family and the range of opportunities that the child might have to gain exposure to reading materials and activities (e.g., number of books in the household). Items in the literacy questionnaire were based on prior measures constructed by Griffin and Morrison (1997) and Sénéchal, LeFevre, Thomas, and Daley (1998).

Procedure—Participation included three visits to the on-campus lab. Participants received a small gift at the end of each visit. During the first visit, the parent was given the training book and instructed to read it aloud to their child, reading the words exactly as written. This reading was videotaped. At this session, a researcher administered the PPVT to the child, and the mother filled out the essentialism and literacy questionnaires. The parent was asked to take the book home and continue reading the book aloud to the child at least once per day until they returned to the lab for their second visit (a 1- to 2-week period), and to keep a reading log to note when readings took place. Dyads were instructed that they could read the book as many times as they wanted to, but to please make sure that they read it at least at least once a day. The second visit occurred 1–2 weeks after the first. The parent-child dyad returned to the lab, bringing with them the training book and their reading log.

This second visit began with a final parent-child reading of the training book to ensure that readings faithfully captured the text as written. The training book was then returned to the researcher and the dyads did not have any access to the training book thereafter. Following the book reading, the parent was taken to a separate testing room (all subsequent participation by parent and child was thus independent of one another). Participants next completed a 4-minute distracter task: the child played with Legos and the parent completed multi-digit multiplication problems. This ensured that information from the most recent book reading was no longer in short-term memory.

Following the distracter task, the parent filled out a paper-and-pencil version of the test task battery as in Study 1, with tasks presented in the same order as in Study 1. The child also completed the same test task battery in the same order, but was administered an oral version by the researcher. The test items and questions were identical to the adult version of the test battery; however, the experimenter showed the child the designated items (printed on laminated cards) on each trial and the child was asked to respond either orally or by pointing. In the child-version of the book-reading task, children were asked to “read” the book to the experimenter, just as they had heard their parent read it. If children did not produce the correct title when “reading” the front cover at the beginning of the task, the experimenter prompted the child by saying the title for him/her and having the child repeat it. This ensured that the child understood the procedure. The child’s test session was videotaped. Following the completion of the test battery, the second visit was concluded.

The third visit took place 2–3 weeks after the second visit. During the third visit, parent and child again received the task battery following the same procedure as in the second visit. Note that neither parent nor child had access to the training book between the second and third visits.

Task Intercorrelations—As in Study 1, we examined the extent to which scores intercorrelated, using Pearson correlations for all analyses. We first examined the extent to which the scores of individual tasks correlated between the two test sessions, for children and parents separately. The scores were highly consistent over the two testing sessions, showing correlations on a given task from Visit-2 to Visit-3 ranging from .38 to .85, *p*s < .01. We next examined intercorrelations within each composite, combined across visits, for children and parents separately. For the Category-Property composite, there were significant intercorrelations for parents among all three tasks (Pearson correlations ranging from .31-.59, *p*s < .05). There were no significant correlations for children. For the Essentialism composite, parents showed significant correlations between explanation (content) and stability (.45, $p = .001$), and between stability and novel induction (.28, $p = .053$), whereas children showed a significant correlation between explanation (content) and inheritance (.41, $p < .01$). For the memory composite, parents showed significant correlations between memory of the noun phrase and memory of the predicate (.37, $p = .01$), and between memory of the predicate and picture memory (.30, $p < .05$). Children showed a significant correlation between memory of the noun phrase and memory of the predicate in the book-reading task (.79, $p < .001$). Although we anticipated that responses within each composite would be significantly correlated for both children and adults, we believe the varying effects could stem from the demands of completing the test task battery for young children (see General Discussion for further discussion).

Results

Two coders scored the explanation (content) task responses, with 20% of the participants' responses (n = 10 dyads responses at Visit 2 and n = 9 dyads responses at Visit 3) coded by both coders to assess reliability. Reliability was high (95%, K = .89). Three coders scored the book-reading task responses, with 33% of the responses (32 participants' responses at Visit 2 and 32 participants' responses at Visit 3) scored by two coders to assess reliability. Reliability was high (99%, K =.97).

We present the results in three phases: First, we present background information on the PPVT and parent questionnaires, with the primary goal of ascertaining whether the scores are equivalent across conditions. Second, to determine if participants in the three conditions received equivalent amounts of training, we present data pertaining to the at-home book readings and length of delay between visits. Third, as in Study 1, we calculated composite scores for each participant for the category-property, essentialism, and memory tasks. We report ANOVAs for the three composite scores, and examine the effects of wording condition, participant age, and testing session (visit) on test performance. Means for all tasks are presented in Figures 4 and 5.

Background information—There were no significant differences across conditions in child PPVT scores (generic M = 111.25 (raw score), 73.56 (percentile); specific-label M = 115.93 (raw score), 79.93 (percentile); no-label M = 112.75 (raw score), 76.38 (percentile)) or parent essentialism questionnaire scores (generic M = 3.58; specific-label M = 3.65; no-label M = 3.59), Fs < 1.0. Therefore, any differences across conditions on our primary measures cannot stem from differences in child vocabulary or parents' essentialism beliefs. There was a significant difference across conditions on the literacy questionnaire, $F(2,44) = 3.78$, $p < .05$. Post-hoc tests indicated that the literacy score in the specific-label condition (M = 36.40) was significantly higher than that in the no-label condition (M = 32.50), $p < .05$. The generic condition (M = 33.19) was intermediate and not significantly different from either of the other two conditions. Therefore, any advantages of the generic wording condition cannot be accounted for by this factor.

Training Data—The length of delay between lab visits and the amount of exposure to the training book did not vary significantly by condition. The average length between visits 1 and 2 was 6.3 days, and the average length between visits 2 and 3 was 15.8 days. Between visits 1 and 2, most children (69%) were read the book daily. On average, parents and children read the books together at home on 6 days between visits 1 and 2, amounting to an average of 7.3 total readings of the book. Additionally, children read the book with either another person or themselves an average of 1.25 times. Thus, altogether, children received an average of 170 target sentences (17 per book × 10 book readings [2 book readings in lab + 8 book readings at home]), before receiving the test battery for the first time. None of these factors varied by condition.

We next looked specifically at the in-lab book-reading sessions to ensure that parents correctly read the text of the training book to their child. Parents read the training book twice to their child in the lab, once during visit 1, and once at the beginning of visit 2. Out of 1,632 occasions (48 dyads × 2 readings × 17 pages), parents deviated from the wording of the text in the book only 96 times (6%), equally split between the first in-lab reading (47%) and the second in-lab reading (53%). Of these, on only 9 occasions (less than 1% of total possible occasions) was the form or content changed in a substantive way. These involved changing the scope of the subject noun phrase ("Zarpie loves to eat tulips" instead of "Zarpies love to eat tulips"; n=1), changing the scope of the predicate ("This zarpie is scared of ladybug" instead of "This zarpie is scared of ladybugs"; n=2), and changing the content of the predicate (e.g., "This zarpie likes to hop over puddles" instead of "This zarpie hops over puddles"; n=6). The remaining deviations from the text were non-substantive (e.g., "Zarpies wiggle their horns when they're happy" instead of "Zarpies wiggle their horns when they are happy"). Overall, then, parents were extremely accurate in conveying the text of the books as written.

Category-property composite—A repeated measures ANOVA was conducted on the category-property composite, with speaker (child, parent) and condition (generic, specific-label, no-label) as between-subjects factors, and testing session (first — visit 2, second — visit 3) as the within-subjects factor. As predicted, there was a significant effect for condition, $F(2,90) = 45.37$, $p < .001$, $\eta^2 = .50$. Scores were higher in the generic condition than in the specific-label or no-label conditions, ps < .001. There was also a condition × age interaction, $F(2,90) = 20.81$, $p < .001$, $\eta^2 = .32$. Planned t-tests indicated that the condition effect was larger for parents than children. Whereas parents showed a significant difference between the generic condition and both the specific-label and no-label conditions, ps < .001, children showed a significant difference between the generic condition and the no-label condition only, $p < .02$. There was also a condition × session interaction, $F(2,90) = 4.45$, $p < .02$, $\eta^2 = .09$, indicating that, in the generic condition only, the category-property link became significantly *stronger* from testing session 1 to testing session 2, $p < .01$. There were no visit effects in the other two conditions. This interaction is illustrated in Figure 6.

There was also a session × age interaction, $F(1,90) = 5.38$, $p < .05$, $\eta^2 = .06$, indicating that the category-property link strengthened from the first to second testing session for children, $p = .01$, but not for parents. As noted earlier, the boost in test scores across sessions occurred in the generic condition only.

We also examined which individual tasks showed an effect of wording condition. Children showed significant condition effects for the explanation (scope) task at both visits ($p \leq .01$, with the generic condition yielding higher scores than the other two conditions), whereas parents showed significant condition effects for all three tasks (categorization, familiar induction, and explanation (scope)) at both visits, ps < .001. For the parents, the generic wording condition was higher than the other two conditions for categorization (Visit 3), familiar induction (both visits), and explanation (scope) (both visits). Additionally, for the categorization task, the generic wording condition was higher than the specific wording condition at Visit 2, and the no-label condition was higher than the specific-label condition (both visits).

Essentialism composite—A repeated measures ANOVA was conducted on the essentialism composite, with speaker (child, parent) and condition (generic, specific-label, no-label) as between-subjects factors, and testing session (first, second) as the within-subjects factor. As predicted, there was a significant effect for condition, $F(2,90) = 9.47$, $p < .001$, $\eta^2 = .17$. Planned t-tests indicate that the generic condition was significantly higher than each of the other two conditions, which did not differ from one another. There were no significant interactions involving age. However, we conducted planned comparisons within each age group separately, to determine whether condition effects on the essentialism composite held for both age groups. These analyses indicated that, whereas parents showed a significant difference between the generic condition and both the specific-label and no-label conditions, ps < .01, children showed a significant difference only between the generic and the no-label conditions, $p < .05$.

We also examined which individual tasks showed effects of wording condition. Children showed significant condition effects for the taxonomic vs. thematic vs. shape task at both visits (generic and specific-label greater than no-label), ps < .01, whereas adults showed significant condition effects for explanation (content) at both visits (generic higher than the other two conditions), *p*s < .005, stability at both visits (generic higher than specific-label at both visits; generic higher than no-label at Visit-2), *p*s < .05, and novel induction at Visit-3 (generic higher than the other two conditions), $p < .02$.

Memory composite—A repeated measures ANOVA was conducted on the memory composite, with speaker (child, parent) and condition (generic, specific-label, no-label) as between-subjects factors, and testing session (first, second) as the within-subjects factor. As predicted, there was no significant effect of condition, $F(2,90) = 1.78$, $p > .17$. However, there was a significant effect of age, $F(1,90) = 3.94$, $p = .05$, $\eta^2 = .04$, indicating that parents (M= 80%) showed better recall than children (M = 72%). Furthermore, there was a significant effect of visit, $F(1,90) = 20.32$, $p < .001$, $\eta^2 = .18$, indicating that memory declined slightly from the first to the second testing session (*M*s = 80% and 72%). This is not surprising, given that there was a longer delay before the second testing session.

We also examined which individual tasks showed significant condition effects. Children showed no significant effects involving wording condition, for any of the three memory tasks, whereas adults showed a significant condition effect for memory of the predicates in the book-reading task at Visit-3 only, with highest recall in the specific-label condition and lowest recall in the generic condition, $p < .05$.

As in Study 1, recall was highly accurate across all three conditions and both age groups, with participants recalling roughly 75% of the relevant information (pictures and text) provided in the books. Importantly, most errors concerned details of the predicates. Memory for the target noun phrase (which was central to the condition manipulation) was uniformly excellent: 82% for the children, and 93% for the parents.

Discussion

The results of this study both replicate and expand on the findings from Study 1. Parents' performance in this study largely mirrored that of the adults in Study 1: both category-property links and essentialism were boosted by hearing generic noun phrases. However, in contrast to Study 1, we obtained a generic advantage only, and no advantage for the specific-label condition. Recall that in Study 1, the specific-label condition led to greater essentialism than the no-label condition. This labeling advantage did not appear for Study 2 adults. A comparison of the means for the adults in Studies 1 and 2 suggests that essentialism scores in the no-label condition are somewhat higher in Study 2. It may be that the more extensive training input in this study, and/or the delay between training and testing strengthened the degree of essentialism – even in the no-label condition.

This study expands on Study 1 in two important ways: it reveals that generic input also influences the construction of a novel concept in preschoolers, and it shows how the effect of generic input on concepts evolves over time. Of particular interest is that preschoolers also show conceptual implications of hearing generic noun phrases. The effects were not as strong or consistent with children as with adults, and were most evident when examining the full composites rather than the individual tasks. We suspect that these age differences are likely due at least in part to demands of the procedure. Participants were presented with a lengthy test battery with many different kinds of questions that required switching from one sort of question to another multiple times within the space of minutes. Furthermore, all test questions were presented with specific (non-generic) wording, thus requiring participants to rely on long-term recall of the book to make use of the wording distinctions. These task demands were likely to have obscured smaller or more subtle wording effects, particularly among children. Another possibility (not mutually exclusive with that of task demands) is that children may truly be less sensitive to wording effects than adults (e.g., Gelman & Bloom, 2007; Hollander, Gelman, & Raman, 2009).

Nonetheless, despite these potential obstacles, children were significantly more likely to link category and property and were significantly more likely to endorse essentialist implications after hearing repeated generic attributions for the new category, as compared to hearing no label for the novel animals. Hearing attributions with specific labels was intermediate, and not significantly different from either the generic wording condition or the no-label condition.

One key question in the current study was whether the effects of generic input on concepts would persist after a delay between training and testing. As noted earlier, both testing points involved a delay. At testing session 1, participants completed a final book reading, but a several-minute distracter task occurred between this reading and the start of the first testing session. The second testing session took place on average more than 2 weeks after the first, and no contact with the book was provided during the entire interval. Importantly, no generic language was used during either testing session. Despite these factors that might seemingly work against any potential effects of generic training, condition differences did not decrease with time, and in the case of category-property links, actually strengthened over time. Specifically, scores for the category-property links increased for the generic wording. (There were no changes over time in the other two wording conditions.) Thus, wording effects are persistent, and only increase as memories for details of the input diminish. We speculate that the strength of participants' generic representations of the zarpie category may be inversely related to the strength of their specific representations of zarpies. Thus, as memory for specific exemplars and/or exemplar-property pairings fade, participants' generic concepts may gain in strength.

Particularly given these cumulative effects of wording, we were interested in examining whether the same effects would obtain when experience with the relevant language was more minimal (even more minimal than the one-time book-reading training in Study 1). Accordingly, Study 3 examined generic, specific-label, or no-label input about zarpies, but with piecemeal and limited rather than extensive input. Specifically, instead of experiencing the training book before testing, participants heard only one relevant sentence immediately before each test question. Again, both preschool children and their parents participated.

Study 3

Methods

Participants—50 preschool children (22 females, 28 males; mean age 52.8 months; age range 48–60.5) and 48 parents (43 females, 5 males) participated. Most participants were White (n = 43 dyads). They received a modest gift for their participation. Participants were randomly assigned to condition (children: Ns = 16, 17, and 17 in the generic, specific-label, and so-label conditions; parents: Ns = 16, 17, and 15 in the generic, specific-label, and no-label conditions). One child and two parents (all from different dyads) completed the study but were excluded because they did not complete the test battery (child) or because of experimenter error in assembling the test battery packets (two parents).

Materials—Materials included test battery booklets (parents) and test cards (children), as in Study 2. There was no training book in this study. Additional materials were the essentialism and literacy questionnaires from Study 2.

Procedure—The same tests as in Studies 1 and 2 were included, with the exception of the memory tasks (picture memory, blank book), as the book was never presented and therefore could not be recalled. The tasks were presented in the same order as in the earlier studies.

Participants did not view the training book. Instead, individual items were presented on each test trial, with relevant information, including pictures and properties that had been provided in the training book. For example, on the familiar induction task, participants first heard the relevant property and saw an accompanying picture concerning the target item (e.g., “Zarpies are afraid of ladybugs” or “This zarpie is afraid of ladybugs” or “This is afraid of ladybugs”), followed by the test questions (“Is this (zarpie) afraid of ladybugs?”). Wording for all tasks is provided in the Appendix.

Note that the familiar induction and novel induction tasks are equivalent in this study because none of the properties were made familiar through training. However, in order to permit comparison with the prior studies, we separate them as in Studies 1 and 2.

Task Intercorrelations—As in Studies 1 and 2, we examined the extent to which scores intercorrelated. For the Category-Property composite, all three tasks intercorrelated for adults: familiar induction and explanation (scope) (.59, $p < .001$), familiar induction and categorization (.42, $p < .01$), and categorization and explanation (scope) (.27, $p = .064$). For children, none of the scores significantly intercorrelated. For the Essentialism composite, adults showed significant correlations between novel categorization and the taxonomic vs. thematic vs. shape task (.34, $p < .02$), and between novel categorization and explanation (content) (.41, $p < .01$). Children displayed a significant correlation between novel categorization and stability (.35, $p < .02$).

Results

Two coders scored participants’ responses in the explanation (content) task, with 20% of the responses ($n = 20$ dyads) coded by both coders to assess reliability. Reliability was high (87%, K = .65).

We present the results in two phases: First, we present background information on the parent questionnaires, to determine whether the conditions are equivalent. We then report separate ANOVAs for the category-property composite and the essentialism composite, to examine the effects of wording condition and participant age. The data are presented in Figures 7 and 8.

Background information—There were no significant condition differences for either the child PPVT scores (*M*s = 73.81, 75.88, and 81.88 for the generic, specific-label, and no-label conditions) or the parent essentialism questionnaire scores (*M*s = 3.53, 3.50, and 3.56 for the generic, specific-label, and no-label conditions), $Fs(2,47) < 1.0$.

Category-property composite—A univariate ANOVA was conducted on the category-property composite, with age (child, adult) and condition (generic, specific-label, no-label) as between-subjects factors. As predicted, there was a significant effect for condition, $F(2,92) = 19.92$, $p < .001$, $\eta^2 = .30$. Scores were higher in the generic condition than in the specific-label or no-label conditions, $ps < .001$. However, there was also a condition × age interaction, $F(2,92) = 22.40$, $p < .001$, $\eta^2 = .33$. Whereas adults showed a significant difference between the generic condition and both the specific-label and no-label conditions, $ps < .001$, the preschool children showed no significant condition effects.

We also examined which individual tasks showed significant effects of wording condition. Children showed no significant condition effects, whereas parents showed significant condition effects for categorization ($p < .001$, with specific-label lower than the other two conditions, and generic trending higher than the no-label condition), familiar induction ($p < .001$, with generic higher than the other two conditions), and explanation scope ($p < .001$, with generic higher than the other two conditions).

Essentialism composite—A univariate ANOVA was conducted on the essentialism composite, with age (child, parent) and condition (generic, specific-label, no-label) as between-subjects factors. The only significant effect was a condition × age interaction, $F(2,92) = 9.22$,

$p < .001$, $\eta^2 = .17$. Planned comparisons indicate that all three conditions are significantly different from one another for the parents (generic greater than both specific-label, $p = .058$, and no-label, $p < .001$; and specific-label greater than no-label, $p < .05$). In contrast, children show a tendency to respond higher in the no-label than generic conditions ($p = .059$) — in other words, the opposite of what was obtained in Study 2.

We also examined which individual tasks showed significant effects of wording condition. Children showed a significant condition effect for inheritance ($p < .02$, with *lowest* scores in the generic wording condition), and parents showed significant condition effects for novel induction, $p < .001$, and the taxonomic vs. thematic vs. shape task, $p < .01$. For the parents, the no-label condition was significantly higher than the other two conditions, and for novel induction, there was a trend for generics to be higher than the specific-label condition.

Discussion

Even with minimal input, adults once again showed effects of wording: generic wording led to tighter category-property links and greater category essentialism compared to both the specific-label condition and the no-label condition. Furthermore, the specific-label condition led to greater category essentialism compared to the no-label condition. This specific-label effect is the same as we saw with adults in Study 1, though it was absent in Study 2. It would appear that with minimal input (as in Studies 1 and 3), simply hearing a label for a category that includes disparate exemplars is sufficient to encourage an essentialist view of the category, as compared to hearing no label for these instances. In contrast, with more extended input and experience with the category (as in Study 2), specific-labeling effects disappear.

In contrast to the adults, children did not show any of the predicted wording effects. Indeed, they showed a tendency to show somewhat higher essentialism scores in the condition in which none of the pictures were labeled. It is surprising that children did not show any wording effects, given prior studies showing that hearing a generic leads to tighter category-property links (Chambers et al., 2008; Hollander et al., 2009). However, the present set of tasks was particularly demanding, as the test battery included many questions and shifted from one task to the other (see Discussion of Study 2). Thus, the lack of a significant wording effect should be interpreted cautiously.

General Discussion

In this section, we first summarize the main contributions and implications of these three experiments, then summarize the limitations and unanticipated findings of the work, and finally suggest directions for future research.

Major contributions and implications

As predicted, the current findings support the hypothesis that generics have special conceptual implications: they strengthen the link between categories and properties and promote an essentialist stance regarding categories. Thus, generics have implications beyond expressing that a particular property extends broadly to the category referred to. Generics are not limited to conveying information about the scope of the properties mentioned in the input. For example, “Zarpies hate ice cream” has implications not only for the link between zarpies and ice cream, but also for the category of zarpies more generally. For both adults and preschool children, generic nouns referring to a novel category exerted significant effects compared to a no-label condition.

The effects of generics in the present studies are striking in three respects. First, the novel category (zarpies) consisted of animals. It has been suggested that animals may be readily essentialized even in absence of input language (Diesendruck, 2003). Generic wording effects may be less powerful when expressing animal categories than social categories, which are more variably interpreted across cultures and individuals. Second, generic information was not available during testing, for Studies 1 and 2, and indeed all the test sentences referred to specific individuals. Thus, in order to make use of the condition differences in Studies 1 and 2, participants had to recall and bring to mind the information provided during the reading of the training books. Third, no input was provided between the first and second testing sessions (in Study 2). Despite the absence of a training book during this period, the effects of extended input were even more powerful over time. In all of these respects, the experiments provided strong tests of the effects of generic language. Nonetheless, clear wording effects obtained, thus suggesting a powerful and unique effect of generics on concepts.

In addition to the generic wording effects, there is also some support for the idea that, at least for adults, labels per se promote an essentialist stance regarding categories. As noted in the Introduction, psychological essentialism includes two separable components: a kind component (that certain categories are structured as natural kinds, with non-obvious, inherent commonalities), and an essence component (that there is an internal part or substance — such as DNA, blood, or even some unknown quality — that is responsible for outward properties that category members share). For adults in Studies 1 and 3, hearing the same label applied to disparate category exemplars led to greater essentialism in this second sense, namely, by fostering a view of zarpies as a natural kind. This finding is consistent with prior work arguing that labels promote taxonomic choices (Markman, 1989; Waxman & R. Gelman, 1986), inductive inferences (Gelman & Markman, 1986; Graham, Kilbreath, & Welder, 2004; Sloutsky & Fisher, 2004; Yamauchi, 2005), kind concepts (Xu, 2005; Yoshida & Smith, 2005), and category stability (Gelman & Heyman, 1999; Reynaert & Gelman, 2007; Walton & Banaji, 2004).

The present findings also address a long-standing debate in the literature, regarding the source of category-label effects. As noted earlier, numerous researchers have found that labels lead to changes in categorization and induction, but why such effects obtain is a matter of much dispute. One view is that category labels (and generic nouns) have the effects they do because they indicate that a category is a richly structured kind, and children draw inferences on the basis of category membership, not merely on perceptual similarity (Gelman & Waxman, 2007). This is the view that we favor here. The breadth of the effects (extending to essentialist reasoning) argues that generics have wide-ranging influences concerning the structure and content of the novel category.

However, similarity and associative learning accounts of concepts reject this view, arguing for lower-level explanations for labeling effects. On one alternative account, when perceptual cues are misleading or distorted, they will not be sufficiently rich to counteract labels that are provided (Jones & Smith, 1993). For example, if children see a simple line drawing that looks like a worm but is labeled “a snake”, this poses a conflict between the label and pre-existing beliefs about snakes. Given this conflict, children may reject the line drawing as overly simplified or misleading, and so rely instead on the label. However, it is difficult to argue that the present study presented distorted similarity cues, as all the instances are structured around a prototype and share a family resemblance (Rosch, 1978); moreover, none were created to resemble animals in an alternative category. If categories are based on perceptual similarity alone, and if children were attempting to use perceptual similarity to guide their responses, then we should have obtained no condition effects. Children should instead have shown the same results across conditions, because the perceptual cues were identical across the three wording conditions.

On another alternative account, children treat words as just another perceptual feature of the display, and make judgments based on the perceived overall similarity (visual and auditory) among items (Sloutsky & Fisher, 2004; Jones & Smith, 1993). For example, two animals will be perceived as more perceptually similar if they are labeled with the same word (e.g., if both are called "zarpies"). On this view, labels influence responses only indirectly, by affecting perceptual similarity judgments. If this were the case, then performance on the category-property composite should be higher for the generic and specific-label conditions than for the no-label condition (depending on how the label is weighted), but no condition differences should emerge between the generic and specific-label conditions for this composite, because they both involve labeling the items with the same name. Moreover, no condition differences whatsoever should emerge for the majority of essentialism tasks, as they involve complex judgments (e.g., providing an explanation for a property; judging stability of a property over time). In contrast to these predictions, we obtained condition differences for both the category-property composite and the essentialism composite, arguing against the idea that labels are construed merely as a perceptual attribute of the relevant items.

We do not mean to suggest, however, that essentialism requires or emerges out of language. There are good reasons to doubt that language creates an essentialist stance in childhood, given prelinguistic underpinnings to essentialism, as well as kind-based reasoning in children who have little exposure to a structured language input (Gelman, 2003; but see Xu, 2005). Relatedly, generics are unlikely to cue essentialism in any sort of automatic process, given that generics can be used to refer to non-essentialized categories (e.g., artifact categories, as in "cups are for drinking" or "chairs are created by humans"). So, hearing a generic (or set of generics) is unlikely to convert just any category into an essentialized kind. Adults (and children) in these studies already know about the structure of animal kinds, and the present findings are probably parasitic on that knowledge. At the same time, it is important to emphasize that the obtained effects do not simply entail attributing properties to categories of a pre-existing structure, since the input language affects precisely how these novel categories are structured. For example, as the present studies demonstrate, hearing generic language leads participants to generalize wholly novel properties from one category member to another, whereas hearing non-generic language does not.

Limitations and unanticipated findings

One unanticipated finding in the current work is that there are developmental differences in the wording effects. Adults consistently showed generic wording effects compared to two key comparison conditions: the specific-label condition as well as the no-label condition. Across the three studies adults showed a clear and replicable wording effect, no matter how tested (minimal or extended input; immediate or delayed testing; in comparison to both the no-label and specific-label conditions). For adults, generic wording confers a clear bias to essentialize a novel category.

Although children overall showed similar patterns, there were developmental differences as well. In contrast to the adults, children showed more fragile wording effects, evident only when input was extensive. Like adults, preschool children distinguished between the maximally informative condition (hearing a novel category label in generic form) and the minimally informative condition (no novel category label, no generic form). Unlike adults, however, children for the most part failed to show more subtle effects. First, children showed no differences between the specific-label condition and the no-label condition, in any of the comparisons. Second, the generic wording condition differed from the specific-label condition only in the second testing visit of Study 2, and not in the first testing visit of Study 2. Third, when the input was minimal (Experiment 3), children showed no sensitivity to either of the linguistic cues (labeling or generics). Altogether, then, children seemed to be more sensitive to the combination of labeling and generic language than to either cue individually, and they seemed more sensitive to the language cues after receiving extensive experience with the picture book. This is perhaps not surprising, given how much more experience adults have with generics. What is perhaps more notable in our findings is that children as young as preschool age showed similar patterns overall as adults.

One open question is why the results were relatively weak with children. As noted earlier, one relevant factor may be that the test questions in Studies 1 and 2 were all presented in specific wording format, potentially diluting the effects of generic wording as well as requiring participants to carry over information that was learned earlier. (In prior research, generic statements immediately preceded the test questions and were unambiguously generic.) This could not account for the weaker effects in Study 3. However, Study 3 presented an extensive set of diverse reasoning tasks to children, which may have been challenging for children to process.

Another factor that may have underestimated the essentialist implications of generics in children is the use of bare plural generics (e.g., “Zarpies hate ice cream”) as opposed to indefinite singular generics (e.g., “A zarpie hates ice cream”). The bare plural form of the generic can be used to express either principled connections (e.g., “dogs are four-legged” because four-leggedness is a natural state for dogs) or statistical connections (e.g., “barns are red” because the majority of barns happen to be red), whereas the indefinite singular more clearly implicates principled connections (Prasada & Dillingham, 2006, 2009). Furthermore, recent work by Cimpian, Meltzer, and Markman (under review) suggests that preschoolers can interpret indefinite singulars generically. The present study did not include the indefinite plural, as we wished to avoid using a form that is potentially ambiguous and could be interpreted as referring to the individual zarpie in the picture. However, this would certainly be important to examine in future research.

One interesting result is that each composite score displayed a distinct pattern of results. As predicted, the memory composite yielded no condition effects whatsoever, for children or adults. Participants were equally attentive to the training input in all three wording conditions. In contrast, both the category-property composite and the essentialism composite showed clear condition effects. For children, both composites showed a generic advantage relative to the no-label condition, with the specific-label condition intermediate but not different from either. For adults, however, the category-property composite and the essentialism composite showed different patterns from one another. For the category-property composite, generic wording yielded much higher scores than both the specific-label and no-label conditions. However, for the essentialism composite, generic wording was higher than the specific-label condition, which in two of the studies was higher than the no-label condition. In other words, on average, specific labeling resembles the no-labeling condition for the category-property link tasks, but tends to be intermediate between generic and no-labeling conditions for the essentialism tasks.

Directions for future research

Despite the developmental differences in performance, Study 2 permits us to conclude that generics do play a role in both children’s and adults’ concept acquisition, and a comparison of Studies 2 and 3 suggests that more extensive experience with generic input appears to have qualitatively different effects for children than less extensive experience. At the same time, however, the studies do not by themselves establish the boundary conditions for the effects of generic language on concepts. That is a question that remains for future research.

Another important open question concerns whether the effects of the generic wording condition are due to hearing generic statements, or to hearing general statements. Although all generics are general, not all general statements are generic. For example, “All zarpies hate ice cream” is a general statement, but not a generic, because it uses a quantifier to express the generalization. Generics and quantifiers differ semantically in important ways (Cimpian et al., in press; Gelman & Bloom, 2007; Leslie, 2007, 2008). For example, generics are often used to express intrinsic properties, whereas quantifiers more strictly express statistical regularities. A comparison of generics and non-generic expressions of generality would therefore be interesting to pursue in future research. We hypothesize that any kind of general statement (generic or quantified) may strengthen category-property links, but that heightened essentialist inferences may be the result of generics only, not other quantifiers. However, even if generics do not differ from quantifiers such as “all” or “most” in their implications, generics are of special importance, because they are so much more commonly found in the input that children receive (Gelman, Hollander, Star, & Heyman, 2000). So, even if other general statements turn out to have similar effects on people’s inferences, they are unlikely to have such effects in actuality, because they are so rarely used.

Thus, language appears to provide an important way of conveying *which* categories are essentialized. It will be important in future research to examine the generality and limits of this effect. For example, it would be valuable to examine wording effects in different domains, as well as for a range of different kinds of predicates. One interesting future direction would be to examine the implications of the present data for understanding how generic language input may contribute to stereotyping of social kinds. Another as-yet-unexplored direction would be to examine the implications of generic language for essentialism in the strong sense (namely, positing an underlying causal force or substance that category members share; see Gelman, 2003, and Introduction of the present paper). As noted at the beginning of this paper, much of what children learn about the world is through social interactions with others, and generics thus may provide a powerful means through which children learn to essentialize. Extensive generic input regarding a novel category fosters an essentialist stance toward that category, and thus may contribute to early childhood essentialism.

Acknowledgments

This research was supported by NSF grant BCS-0817128 and NICHD grant R01 HD36043 to Gelman. We are grateful to all the children, parents, and college students who participated in the research. We thank Erin Boyle, Melissa Herrett, Anna Koslowski, Amanda Markowitz, and Sarah Stilwell for help with testing, transcription, and coding, and Dave Kush for creating the drawings that were used in the research.

References

Ahn W, Kalish C, Gelman SA, Medin DL, Luhmann C, Atran S, Coley DJ, Shafto P. Why essences are essential in the psychology of concepts. Cognition 2001;82:59–69. [PubMed: 11672705]

Bowerman, M.; Levinson, SC. Language acquisition and conceptual development. New York: Cambridge University Press; 2001.

Brandone AC, Gelman SA. Differences in preschoolers’ and adults’ use of generics about animals and artifacts: A window onto a conceptual divide. Cognition 2009;110:1–22. [PubMed: 19046742]

Chambers CG, Graham SA, Turner JN. When hearsay trumps evidence: How generic language guides preschoolers' inferences about unfamiliar things. Language and Cognitive Processes 2008;23:1–18.

Cimpian A, Arce HC, Markman EM, Dweck CS. Subtle linguistic cues affect children’s motivation. Psychological Science 2007;18:314–316. [PubMed: 17470255]

Cimpian A, Gelman SA, Brandone AC. Theory-based considerations influence the interpretation of generic sentences. Language and Cognitive Processes. in press.

Cimpian A, Markman EM. Preschool children’s use of cues to generic meaning. Cognition 2008;107:19–53. [PubMed: 17765216]

Cimpian A, Markman EM. Information learned from generic language becomes central to children’s biological concepts: Evidence from their open-ended explanations. Cognition. in press.

Cimpian A, Meltzer TJ, Markman EM. Preschoolers' use of morphosyntactic cues to identify generic sentences: Indefinite singular noun phrases, tense, and aspect. under review.

Diesendruck G. Categories for names or names for categories? The interplay between domain-specific conceptual structure and language. Language and Cognitive Processes 2003;18:759–787.

Fulkerson AL, Waxman SR. Words (but not tones) facilitate object categorization: Evidence from 6- and 12-month-olds. Cognition 2007;105:218–228. [PubMed: 17064677]

Gelman, SA. The essential child: Origins of essentialism in everyday thought. New York: Oxford University Press; 2003.

Gelman SA. Learning from others: Children's construction of concepts. Annual Review of Psychology 2009;60:115–140.

Gelman SA, Bloom P. Developmental changes in the understanding of generics. Cognition 2007;105:166–183. [PubMed: 17094957]

Gelman SA, Coley JD, Rosengren K, Hartman E, Pappas A. Beyond labeling: The role of maternal input in the acquisition of richly-structured categories. Monographs of the Society for Research in Child Development 1998;63(1) Serial No. 253.

Gelman SA, Goetz PJ, Sarnecka BS, Flukes J. Generic language in parent-child conversations. Language Learning and Development 2008;4:1–31.

Gelman SA, Heyman GD. Carrot-eaters and creature-believers: The effects of lexicalization on children's inferences about social categories. Psychological Science 1999;10:489–493.

Gelman SA, Heyman GD, Legare CH. Developmental changes in the coherence of essentialist beliefs about psychological characteristics. Child Development 2007;78:757–774. [PubMed: 17517003]

Gelman, SA.; Hollander, M.; Star, J.; Heyman, GD. The role of language in the construction of kinds. In: Medin, D., editor. Psychology of learning and motivation. Vol. 39. New York: Academic Press; 2000. p. 201-263.

Gelman SA, Markman EM. Categories and induction in young children. Cognition 1986;23:183–209. [PubMed: 3791915]

Gelman SA, Raman L. Preschool children use linguistic form class and pragmatic cues to interpret generics. Child Development 2003;24:308–325. [PubMed: 12625452]

Gelman SA, Raman L. This cat has nine lives? Children's memory for genericity in language. Developmental Psychology 2007;43:1256–1268. [PubMed: 17723049]

Gelman SA, Raman L, Gentner D. Effects of language and similarity on comparison processing. Language Learning and Development. in press.

Gelman SA, Star J, Flukes J. Children's use of generics in inductive inferences. Journal of Cognition and Development 2002;3:179–199.

Gelman SA, Waxman SR. Looking beyond looks: Comments on Sloutsky, Kloos, and Fisher, "When looks are everything: Appearance similarity versus kind information in early induction". Psychological Science 2007;18:554–555. [PubMed: 17576270]

Gentner, D.; Goldin-Meadow, S. Language in mind: Advances in the study of language and thought. Cambridge, MA: MIT Press; 2003.

Gergely, G.; Csibra, G. Sylvia's recipe: The role of imitation and pedagogy in the transmission of human culture. In: Enfield, NJ.; Levinson, SC., editors. Roots of human sociality: Culture, cognition, and human interaction. Oxford: Berg Publishers; 2006. p. 229-255.

Goldin-Meadow S, Gelman SA, Mylander C. Expressing generic concepts with and without a language model. Cognition 2005;96:109–126. [PubMed: 15925572]

Graham S, Kilbreath C, Welder A. Thirteen-month-olds rely on shared labels and shape similarity for inductive inferences. Child Development 2004;75:409–427. [PubMed: 15056196]

Griffin EA, Morrison FJ. The unique contribution of home literacy environment to differences in early literacy skills. Early Child Development and Care 1997;127–128:233–243.

Harris PL, Koenig MA. Trust in testimony: How children learn about science and religion. Child Development 2006;77:505–24. [PubMed: 16686784]

Haslam N, Rothschild L, Ernst D. Essentialist beliefs about social categories. British Journal of Social Psychology 2000;39:113–127. [PubMed: 10774531]

Hollander MA, Gelman SA, Raman L. Generic language and judgments about category membership: Can generics highlight properties as central? Language and Cognitive Processes 2009;24:481–505.

Hollander MA, Gelman SA, Star J. Children’s interpretation of generic noun phrases. Developmental Psychology 2002;38:883–894. [PubMed: 12428701]

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

Jones S, Smith L. The place of perception in children's concepts. Cognitive Development 1993;8:113–139.

Kinzler KD, Shutts K, DeJesus J, Spelke ES. Accent trumps race in guiding children’s social preferences. Social Cognition. in press.

Leslie SJ. Generics: cognition and acquisition. The Philosophical Review 2008;117:1–49.

Markman, EM. Categorization and naming in children: Problems in induction. Cambridge: Bradford Book/MIT Press; 1989.

Medin D. Concepts and conceptual structure. American Psychologist 1989;44:1469–81. [PubMed: 2690699]

Prasada S. Acquiring generic knowledge. Trends in Cognitive Sciences 2000;4:66–72. [PubMed: 10652524]

Prasada S, Dillingham EM. Principled and statistical connections in common sense conception. Cognition 2006;99:73–112. [PubMed: 16443448]

Prasada S, Dillingham EM. Representation of principled connections: A window onto the formal aspect of common sense conception. Cognitive Science 2009;33:401–448.

Reynaert CC, Gelman SA. The influence of language form and conventional wording on judgments of illness. Journal of Psycholinguistic Research 2007;36:273–295. [PubMed: 17273932]

Rhodes M, Gelman SA. A developmental examination of the conceptual structure of animal, artifact, and human social categories. Cognitive Psychology. in press.

Rosch, E. Principles of categorization. In: Rosch, E.; Lloyd, BB., editors. Cognition and categorization. Hillsdale, NJ: Erlbaum; 1978. p. 27-48.

Sénéchal M, LeFevre J, Thomas EM, Daley K. Differential effects of home literacy experiences on the development of oral and written language. Reading Research Quarterly 1998;33:96–116.

Sloutsky VM, Fisher AV. Induction and categorization in young children: A similarity-based model. Journal of Experimental Psychology: General 2004;133:166–188. [PubMed: 15149249]

Strevens M. The essentialist aspect of naive theories. Cognition 2000;74:149–175. [PubMed: 10617780]

Tabachnick, BG.; Fidell, LS. Using multivariate statistics. Philadelphia, PA: Harper & Row; 1989.

Vygotsky, LS. Thought and language. Cambridge, MA: MIT Press; 1934/1962.

Walton GM, Banaji MR. Being what you say: The effect of linguistic labels on attitudes. Social Cognition 2004;22:193–213.

Ward T, Vela E, Hass S. Children and adults learn family-resemblance categories analytically. Child Development 1990;61:593–605. [PubMed: 2364735]

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, DG.; Waxman, SR., editors. Weaving a lexicon. Cambridge: MIT Press; 2004. p. 295-335.

Waxman SR. Names will never hurt me? Naming and the development of racial and gender categories in preschool-aged children. European Journal of Social Psychology. in press.

Waxman SR, Gelman R. Preschoolers' use of superordinate relations in classification and language. Cognitive Development 1986;1:139–156.

Waxman SR, Markow DB. Words as invitations to form categories: Evidence from 12-month-old infants. Cognitive Psychology 1995;29:257–302. [PubMed: 8556847]

Wierzbicka A. The universality of taxonomic categorization and the indispensability of the concept 'kind'. Rivista di Linguistica 1994;6:347–364.

Xu, F. Categories, kinds, and object individuation in infancy. In: Gershkoff-Stowe, L.; Rakison, D., editors. Building object categories in developmental time: Papers from the 32nd Carnegie Symposium on Cognition. Mahwah, NJ: Erlbaum; 2005. p. 63-89.

Yamauchi T. Labeling bias and categorical induction: Generative aspects of category information. Journal of Experimental Psychology: Learning, Memory and Cognition 2005;31:538–553.

Yoshida H, Smith L. Linguistic cues enhance the learning of perceptual cues. Psychological Science 2005;16:90–95. [PubMed: 15686573]

Appendix A Test Battery, Studies 1, 2, and 3. Note: Each task is illustrated with one sample item

Categorization

Studies 1 & 2

- *Generic & specific-label*: "Remember this zarpie? Can you find another zarpie?" [two choices]
- *No-label*: "Remember this?" Can you find another one?" [two choices]

Study 3

- *Generic*: "Look at this zarpie. Zarpies have a star behind their horns. Can you find another zarpie?" [two choices]
- *Specific-label*: "Look at this zarpie. This zarpie has a star behind its horns. Can you find another zarpie?" [two choices]
- *No-label*: "Look at this. This has a star behind its horns. Can you find another one?" [two choices]

Familiar Induction

Studies 1 & 2

- *Generic & specific-label*: "Remember this zarpie? This zarpie hates ice cream. Look at this zarpie [test item]. Do you think that this zarpie hates ice cream, like this zarpie [training item]?"
- *No-label*: "Remember this? This hates ice cream. Look at this [test item]. Do you think that this hates ice cream, like this [training item]?"

Study 3

- *Generic*: "Look at this zarpie. Zarpies hate ice cream. Look at this zarpie [test item]. Do you think that this zarpie hates ice cream?"
- *Specific-label*: "Look at this zarpie. This zarpie hates ice cream. Look at this zarpie [test item]. Do you think that this zarpie hates ice cream?"
- *No-label*: "Look at this. This hates ice cream. Look at this [test item]. Do you think that this hates ice cream?"

Explanations

Studies 1 & 2

- *Generic, specific-label, and no-label:* "Why is this chasing a shadow?"

Study 3

- *Generic*: "Look at this zarpie. Zarpies chase shadows. Why is this chasing a shadow?"

- *Specific-label*: "Look at this zarpie. This zarpie is chasing a shadow. Why is this chasing a shadow?"
- *No-label*: "Look at this. This is chasing a shadow. Why is this chasing a shadow?"

Stability

Studies 1 & 2

- *Generic and specific-label*: "Look at this zarpie. This zarpie hops over puddles. You know what else? This zarpie is 8 years old. (a) Do you think it hopped over puddles when it was 4 years old? (b) Do you think it will always hop over puddles?"
- *No-label*: "Look at this. This hops over puddles. You know what else? This is 8 years old. (a) Do you think it hopped over puddles when it was 4 years old? (b) Do you think it will always hop over puddles?"

Study 3

- *Generic*: "Look at this zarpie. Zarpies hop over puddles. You know what else? This zarpie is 8 years old. (a) Do you think it hopped over puddles when it was 4 years old? (b) Do you think it will always hop over puddles?"
- *Specific-label*: "Look at this zarpie. This zarpie hops over puddles. You know what else? This zarpie is 8 years old. (a) Do you think it hopped over puddles when it was 4 years old? (b) Do you think it will always hop over puddles?"
- *No-label*: "Look at this. This hops over puddles. You know what else? This is 8 years old. (a) Do you think it hopped over puddles when it was 4 years old? (b) Do you think it will always hop over puddles?"

Inheritance

Studies 1 & 2

- *Generic and specific-label*: "Look at this zarpie. She wiggles her horns when she is happy. And look at this dog. She wags her tail when she is happy. One day this zarpie had a baby. That means that the baby came out of her tummy. But right after the baby was born, it went to live with this dog. The dog took care of the baby. She played with the baby, fed the baby, and loved the baby. The baby grew up with this dog and never saw this zarpie again. [Memory check questions for children: Whose tummy did the baby come out of? And who took care of the baby?] Now the baby is 6 years old. Do you think that the baby wiggles its horns when it is happy, like this zarpie, or wags its tail when it is happy, like this dog?"
- *No-label*: "Look at this one. She wiggles her horns when she is happy. And look at this dog. She wags her tail when she is happy. One day this one had a baby…. [identical to generic and specific-label conditions, until:] Do you think that the baby wiggles its horns when it is happy, like this one, or wags its tail when it is happy, like this dog?"

Study 3

- *Generic*: "Look at this zarpie. Zarpies wiggle their horns when they are happy. And look at this dog. Dogs wag their tails when they are happy. One day this zarpie had a baby…" [remainder identical to generic condition of Studies 1 and 2]

- *Specific-label*: "Look at this zarpie. This zarpie wiggles its horns when it is happy. And look at this dog. This dog wags its tail when it is happy…." [remainder identical to generic condition]
- *No-label* [same as no-label, Studies 1 & 2]

Novel Induction

Studies 1 & 2

- *Generic and specific-label*: "Look at this zarpie. This zarpie makes a buzzing sound when it is angry. Look at this zarpie [test item]. Do you think that this zarpie makes a buzzing sound when it is angry like this zarpie [target]?"
- *No-label*: "Look at this. This makes a buzzing sound when it is angry. Look at this [test item]. Do you think that this makes a buzzing sound when it is angry like this [target]?"

Study 3

- *Generic*: "Look at this zarpie. Zarpies make a buzzing sound when they are angry. Look at this zarpie [test item]. Do you think that this zarpie makes a buzzing sound when it is angry?"
- *Specific-label*: "Look at this zarpie. This zarpie makes a buzzing sound when it is angry. Look at this zarpie [test item]. Do you think that this zarpie makes a buzzing sound when it is angry?"
- *No-label*: "Look at this. This makes a buzzing sound when it is angry. Look at this [test item]. Do you think that this makes a buzzing sound when it is angry?"

Taxonomic vs. Thematic vs. Shape

Studies 1 & 2

- *Generic and specific-label*: "See this zarpie? This zarpie is a *blicket*. Can you find another blicket?" [three choice cards are presented: a taxonomic match, a thematic match, and a shape match.]
- *No-label*: "See this? Can you find another one?" [same three choice cards]

Study 3

- *Generic*: "Look at this zarpie. [training item] Zarpies sleep in tall trees. You know what else? This zarpie is a *toma*. Can you find another toma?" [same three choice cards]
- *Specific-label*: "Look at this zarpie. [training item] This zarpie sleeps in a tall tree. You know what else? This zarpie is a *toma*. Can you find another toma?" [same three choice cards]
- *No-label*: "Look at this. [training item] This sleeps in a tall tree. Can you find another one?" [same three choice cards]

Figure 1.
Sample zarpies

NIH-PA Author Manuscript
NIH-PA Author Manuscript

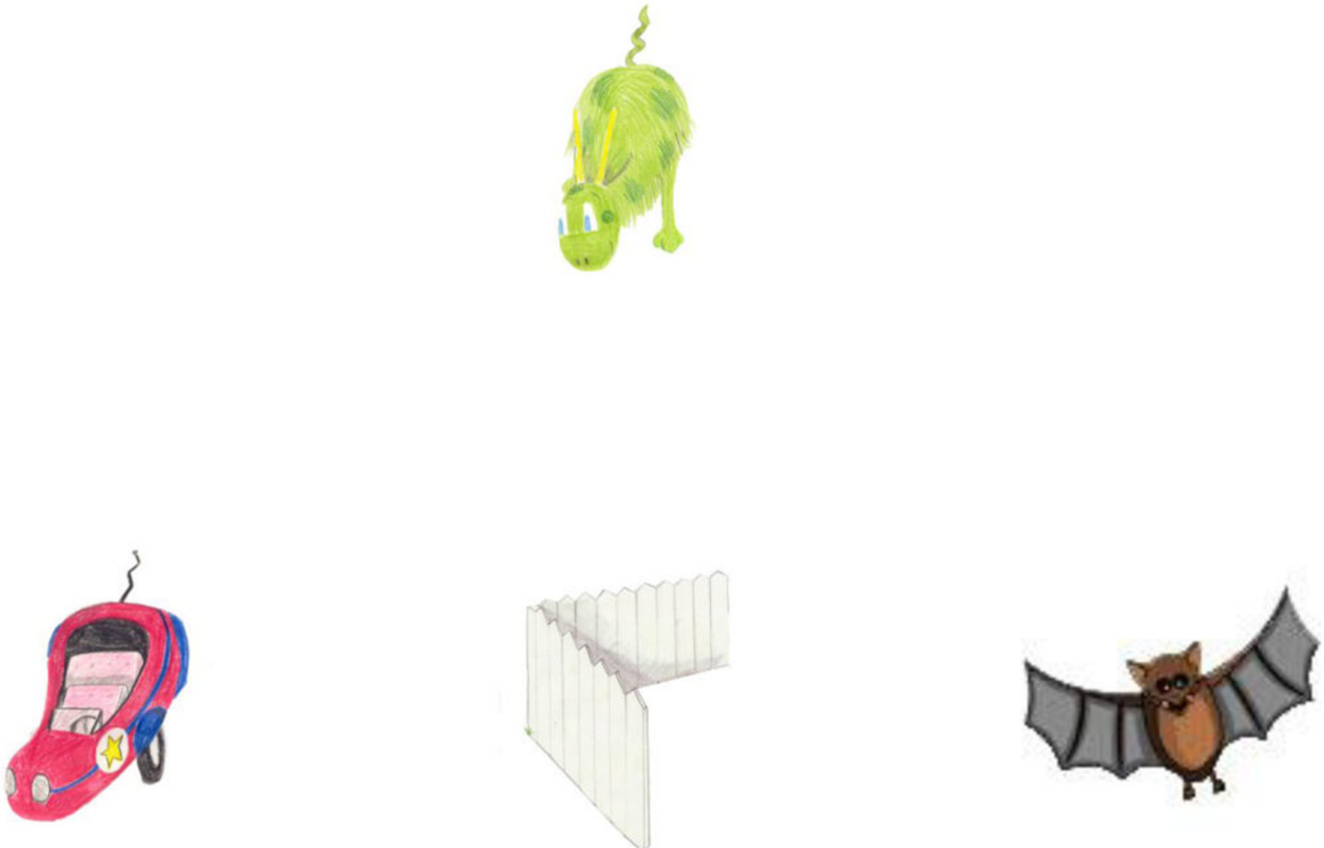

Figure 2.
Sample set from the taxonomic/thematic/shape task (shape choice: race- car; thematic choice: fence; taxonomic choice: bat).

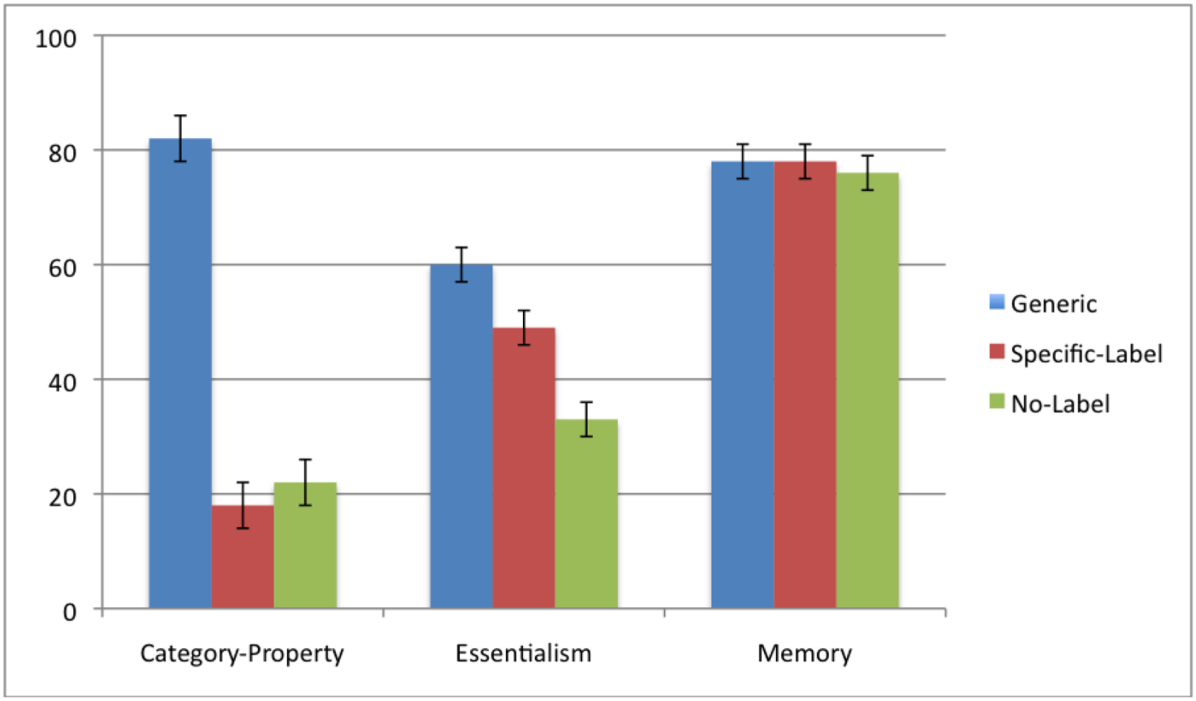

Figure 3.
Study 1 – College Students, mean responses as a function of task and wording condition. Note: Scores were converted to a uniform 0–100 scale by dividing the obtained scores by the total possible score and multiplying by 100. However, statistical analyses are based on mean summed scores (rather than mean percentages), as the analyses are conducted for each composite separately.

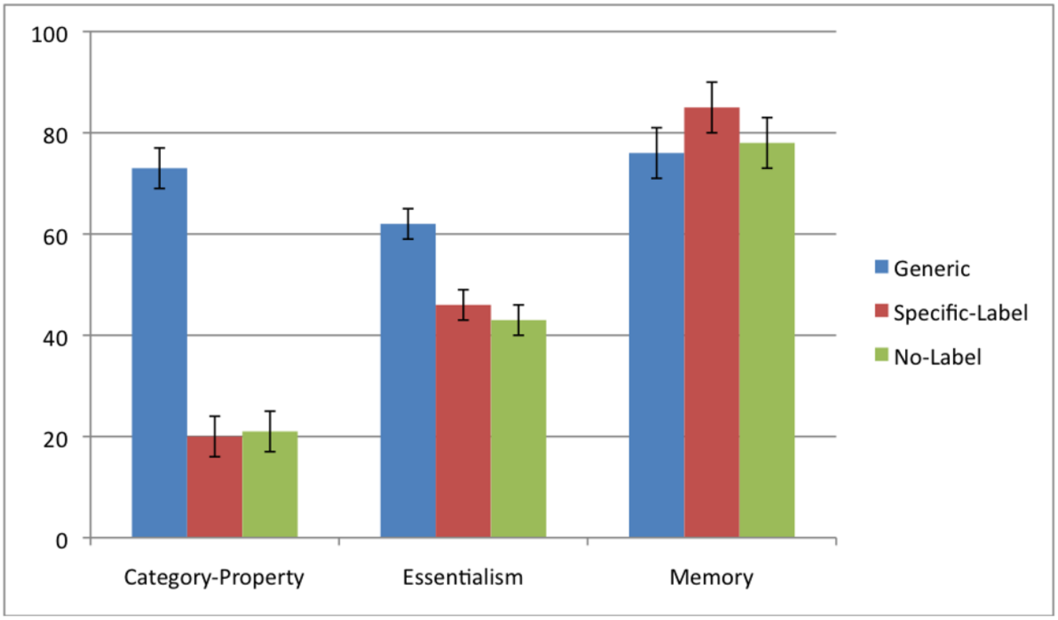

Figure 4.
Study 2, Parents, mean responses as a function of task and wording condition. Note: Scores were converted to a uniform 0–100 scale by dividing the obtained scores by the total possible score and multiplying by 100.

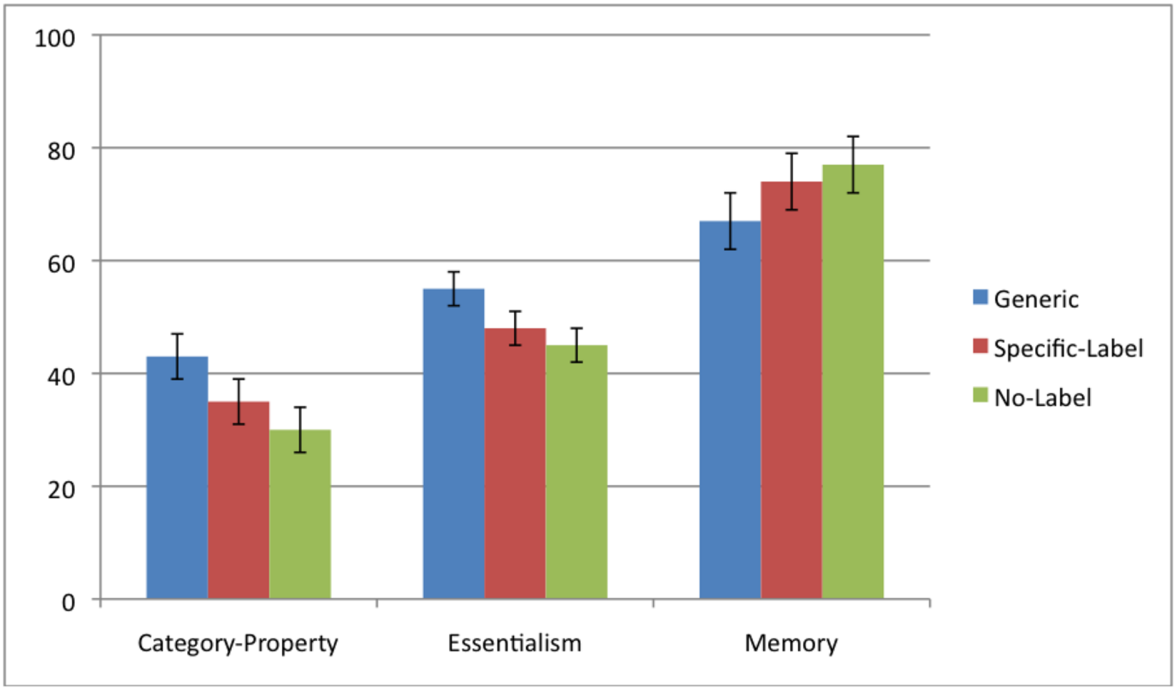

Figure 5.
Study 2, Children, mean responses as a function of task and wording condition. Note: Scores were converted to a uniform 0–100 scale by dividing the obtained scores by the total possible score and multiplying by 100.

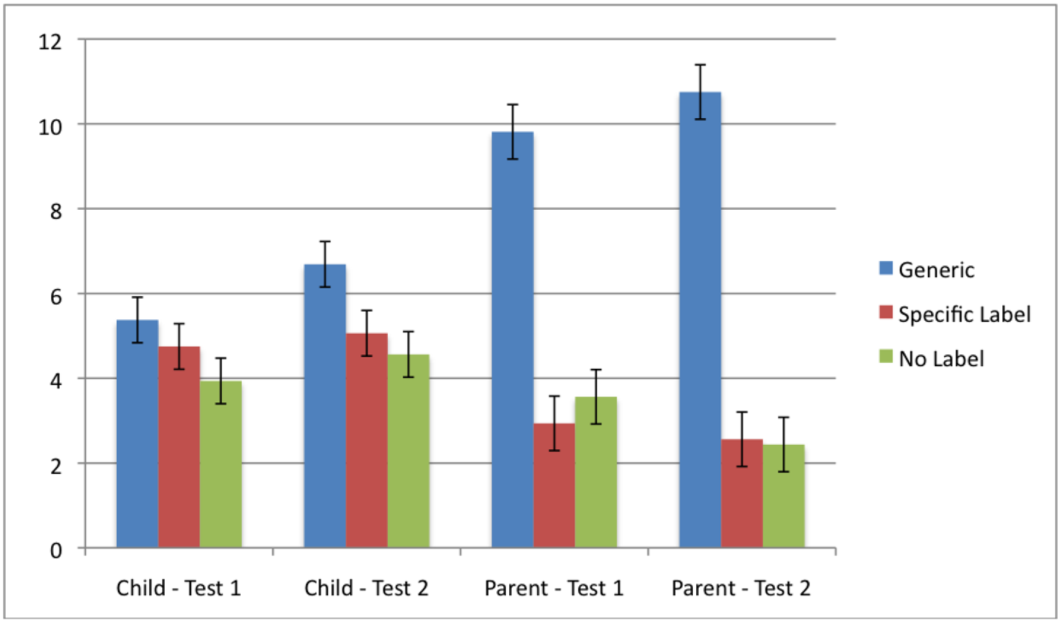

Figure 6.
Study 2, mean category-property composite, as a function of visit, condition, and age.

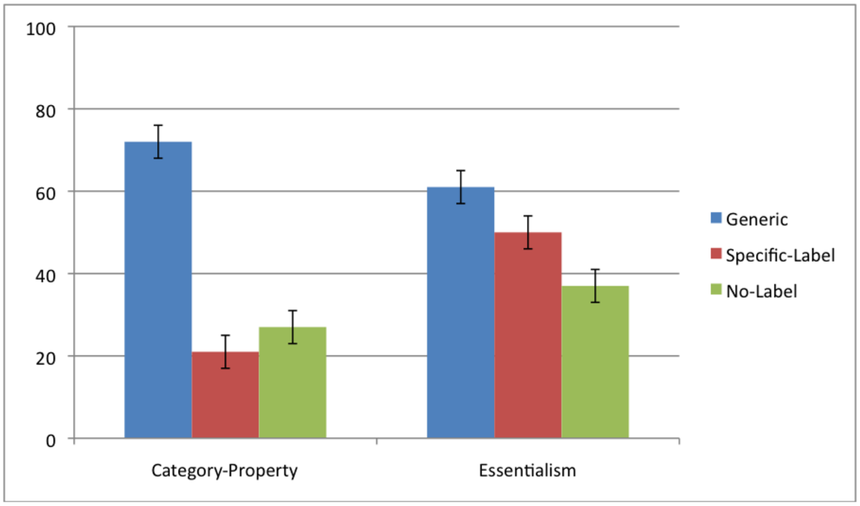

Figure 7.

Study 3, Parents, mean responses as a function of task and wording condition. Note: Scores were converted to a uniform 0–100 scale by dividing the obtained scores by the total possible score and multiplying by 100.

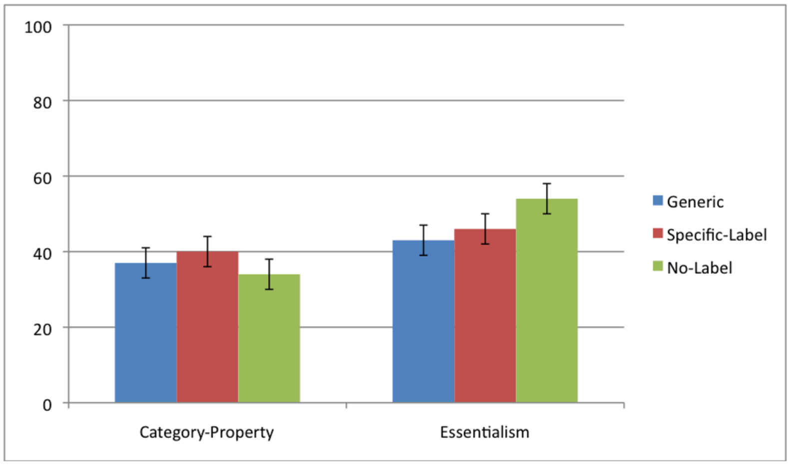

Figure 8.
Study 3, Preschoolers, mean responses as a function of task and wording condition. Note: Scores were converted to a uniform 0–100 scale by dividing the obtained scores by the total possible score and multiplying by 100.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 1

Text of training book.

PAGE	TEXT (IN GENERIC FORM)
(Cover)	All about zarpies!
1	Zarpies love to eat tulips.
2	Zarpies have stripes on the bottom of their feet.
3	Zarpies can bounce a ball on their backs.
4	Zarpies like to sing.
5	Zarpies hide behind fences.
6	Zarpies wiggle their horns when they are happy.
7	Zarpies have spots on their tongues.
8	Zarpies hop over puddles.
9	Zarpies hate walking in snow.
10	Zarpies have a star behind their horns.
11	Zarpies can flip in the air.
12	Zarpies are scared of ladybugs.
13	Baby zarpies have orange fur.
14	Zarpies stand on their back legs when they are surprised.
15	Zarpies chase shadows.
16	Zarpies hate ice cream.
17	Zarpies sleep in tall trees.

Table 2

Feature structure of zarpies in picture book

Page	Fur	Leg	Horn	Tail	Snout	Color	Pattern
Cover	1	3	2	1	1	1	3
Cover	3	1	1	2	1	2	2
Cover	1	1	1	3	2	3	1
1	1	1	1	2	3	3	2
2	3	1	1	1	2	2	1
3	1	3	1	2	1	3	3
4	2	1	3	1	1	2	1
5	1	2	3	1	1	2	3
6	1	1	2	3	1	1	1
7	1	3	1	1	2	3	2
8	2	1	1	1	3	2	3
9	1	1	2	1	3	3	1
10	1	1	3	2	1	1	3
11	3	2	1	1	1	2	2
12	1	1	3	1	2	1	2
13	1	2	1	3	1	4	1
14	3	1	2	1	1	1	2
15	2	3	1	1	1	3	3
16	1	2	1	1	3	2	2
17	2	1	1	3	1	1	1

Key	1	2	3
fur type	long*	short	curly
leg/feet size	short fat legs, big feet*	short thin legs, small feet	long thin legs small feet
horn	curved*	little, rounded at top	long, pointy, wavy
tail shape	pig-like*	horse-like	monkey-like
snout shape	medium*	wide	narrow
fur color	blue	green	purple

Key	1	2	3
pattern	solid	striped	spotted

* indicates prototypical value for this dimension

Table 3

Test battery

Task	**# of items**
CATEGORY-PROPERTY COMPOSITE	
Categorization	4
Familiar Induction	6
Explanation-scope	4
ESSENTIALISM COMPOSITE	
Explanation-content	4
Stability	6
Inheritance	4
Novel Induction	6
Taxonomic vs. Thematic vs. Shape	4
MEMORY COMPOSITE	
Picture Memory	4
No-label Book – Subject Nouns	17
No-label Book – Predicates	58
BACKGROUND	
Essentialism Questionnaire (adults/parents only)	
Literacy Questionnaire (parents in Studies 2 and 3 only)	
PPVT (children only)	

Table 4

Sample coding of explanation responses

Response	Scope	Content
Because it is cold and he has no shoes on.	Non-generic	Non-intrinsic
Because it was scared of a dog and they wanna eat him.	Non-generic	Non-intrinsic
‘Cause it’s not very smart.	Non-generic	Intrinsic
Because it doesn’t have any fur on its toes.	Non-generic	Intrinsic
Because they’re angry.	Generic	Non-intrinsic
Because zarpies chase shadows; that’s what they do.	Generic	Intrinsic
They were born to do it.	Generic	Intrinsic