

# Is the most representative skunk the average or the stinkiest? Developmental changes in representations of biological categories<sup>☆</sup>



Emily Foster-Hanson<sup>\*</sup>, Marjorie Rhodes

Department of Psychology, New York University, USA

## ARTICLE INFO

### Keywords:

Conceptual development  
Biological reasoning  
Ideals

## ABSTRACT

People often think of categories in terms of their most representative examples (e.g., robin for BIRD). Thus, determining which exemplars are most representative is a fundamental cognitive process that shapes how people use concepts to navigate the world. The present studies ( $N = 669$ ; ages 5 years – adulthood) revealed developmental change in this important component of cognition. Studies 1–2 found that young children view exemplars with extreme values of characteristic features (e.g., the very fastest cheetah) as most representative of familiar biological categories; the tendency to view average exemplars in this manner (e.g., the average-speeded cheetah) emerged slowly across age. Study 3 examined the mechanisms underlying these judgments, and found that participants of all ages viewed extreme exemplars as representative of novel animal categories when they learned that the variable features fulfilled category-specific adaptive needs, but not otherwise. Implications for developmental changes in conceptual structure and biological reasoning are discussed.

## 1. Introduction

As people use categories to navigate the world, category members that people view as clear representatives of their kinds (e.g., a robin for the category BIRD) take on more powerful roles than those that are viewed as more peripheral (e.g., a penguin; Murphy, 2002; Rips, 1975; Rips, Shoben, & Smith, 1973; Rosch, 1973). This graded structure of natural categories, first identified by Rosch (1973), shapes a wide range of cognitive processes. For example, representative examples are learned more easily (Bjorklund & Thompson, 1983; Rosch, Simpson, & Miller, 1976), acquired earlier in development (Mervis & Pani, 1980), verified more quickly as category members (Murphy & Brownell, 1985; Rosch, 1973), brought to mind more readily (Anglin, 1986; Rosner & Hayes, 1977), and are thought to provide more generalizable information (Osherson, Smith, Wilkie, López, & Shafir, 1990; Rips, 1975). Thus, determining how people evaluate representativeness is fundamental to understanding the structure and function of concepts in everyday cognition. The goal of the present research is to examine how this important feature of cognition changes across development.

Adults consider several criteria to evaluate representativeness (often referred to as exemplar typicality, Rosch, 1973; Rosch & Mervis, 1975). An ongoing debate in this area is the extent to which people view idealized or average exemplars as better representatives of their kinds (Davis & Love, 2010; Kim & Murphy, 2011; Levering & Kurtz, 2006). Barsalou (1985) proposed that for

<sup>☆</sup> This research was supported by National Science Foundation grant BCS 1729540 (to M.R.) and by the Eunice Kennedy Shriver National Institute of Child Health & Human Development of the National Institutes of Health under Award Number F31HD093431 (to E.F.H.). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

<sup>\*</sup> Corresponding author at: 6 Washington Place, New York, NY 10003-6634, USA.

E-mail address: [emily.fosterhanson@nyu.edu](mailto:emily.fosterhanson@nyu.edu) (E. Foster-Hanson).

some categories, particularly those with clear goals, people often view idealized exemplars as representative (e.g., that a person who donates all of their income to charity better represents the category of “generous person” than someone who donates 1%; see also Borkenau, 1990; Burnett, Medin, Ross, & Blok, 2005; Chaplin, John, & Goldberg, 1988; Lynch, Coley, & Medin, 2000; Read, Jones, & Miller, 1990; Rein, Goldwater, & Markman, 2010). Others have proposed that ideal exemplars play a more limited role in category representations, however, both because many categories do not have clear goals and because methodological features could sometimes lead adults to appear to view ideal exemplars as representative when they really do not (Kim & Murphy, 2011). For example, Kim and Murphy (2011) suggest that adults sometimes misunderstand requests to pick “the best example” as requests to pick “the best one”, and therefore think they are directly being asked to pick idealized rather than representative examples. In a series of studies run with this concern in mind, Kim and Murphy found no evidence that adults viewed ideal exemplars as representative; instead, adults robustly viewed average exemplars in this manner, even for goal-based categories. A tendency to view average, instead of idealized, exemplars as informative is consistent with Rosch and Mervis (1975) and a number of other studies concluding that adults evaluate representativeness not by considering category ideals, but by assessing the extent to which exemplars are similar to other category members and dissimilar to non-members on key features (Ameel & Storms, 2006; Davis & Love, 2010; Goldstone, Steyvers, & Rogosky, 2003; Kim & Murphy, 2011; Medin, Wattenmaker, & Hampson, 1987).

Concepts of natural categories are organized around representative examples from the earliest stages of acquisition (Dewar & Xu, 2010; Mervis & Pani, 1980). Thus, whether the processes that underlie beliefs about category representativeness are consistent or change across childhood is fundamental to our understanding of conceptual development. Yet very little previous work has considered this issue (see Carey, 1985; Medin, Waxman, Woodring, & Washinawatok, 2010). To do so, here we examined representations of the biological world across childhood development, by testing the extent to which children (ages 5–10) and adults view average or idealized exemplars of animal species as most representative and informative regarding their kinds.

To explore the role of ideals in representations of the biological world, we defined ideals not in terms of category goals, but as those exemplars that have extreme values on properties that people view as a normative component of category membership. To illustrate, adults (Prasada & Dillingham, 2006, 2009) and young children (Haward, Wagner, Carey, & Prasada, 2018) think that cheetahs *should* run fast, and that there is something wrong with one that does not; thus, here we defined idealized category members as those with extreme values of these characteristic properties (e.g., the very fastest cheetah in the world). In contrast with previous work (Barsalou, 1985; Burnett et al., 2005; Kim & Murphy, 2011; Lynch et al., 2000), this operationalization of ideals is separable from what might be considered ideal from a human perspective. For example, “stinkiness” is a normative property of being a skunk, in that people agree that skunks are supposed to be stinky, and that there is something odd about a skunk that is not (Haward et al., 2018; Prasada & Dillingham, 2006, 2009). On the definition of ideals that we used in the present paper, therefore, the ideal skunk is one that is extra stinky, even though an extra-stinky skunk would not fulfill (and might be counter to) human goals.

Here we tested whether idealized exemplars—defined in this manner—might play a particularly strong role in the concepts of young children, with a shift to more average exemplars across development. There were at least three reasons for predicting such a shift. First, across various tasks and domains, young children often conflate their ideas about how things *are* with how they *should* be (e.g., assuming that what they observe reflects prescriptive norms, even based on very limited evidence; Kalish, 1998; Rakoczy & Schmidt, 2013; Roberts, Ho, & Gelman, 2017). Young children are also more likely to prioritize prescriptive norms over other information to infer what category members will be like (Kalish & Lawson, 2008; Kalish & Shiverick, 2004). Preschoolers even conflate judgments of probability, possibility, and permissibility (e.g., saying it would be *immoral* for someone to float in the air and *impossible* for someone to steal or tell a lie; Shtulman & Phillips, 2018; see also Phillips & Cushman, 2017; Shtulman & Carey, 2007; Tisak & Turiel, 1988). Given this tendency to conflate descriptive and prescriptive information (and to readily adopt prescriptive views), we thought that young children might value exemplars that they view as best exemplifying category norms, even if such exemplars are unusual (e.g., to focus on the belief that cheetahs *should* run fast over information indicating that actual cheetahs vary in their running speed, and thus to view the very fastest cheetah as better representing cheetahs than an average one).

Second, in addition to young children's general tendency to emphasize prescriptive beliefs, certain features specific to children's biological reasoning and causal beliefs may make them more likely to focus on category ideals in this particular domain. For example, we suspected that the tendency to engage in teleological reasoning (e.g., saying that an unfamiliar animal has smooth skin “*so that* it could move easily through the water”; Kelemen, 1999, 2003; Kelemen, Callanan, Casler, & Pérez-Granados, 2005), which is particularly prevalent in early childhood, would lead children to hold prescriptive beliefs about features that they view as offering a functional or adaptive benefit for the organism.

Third, young children are also especially likely to overlook variation among members of biological categories (Diesendruck & Gelman, 1999; Emmons & Kelemen, 2015; Shtulman & Schultz, 2008). For instance, adults recognize the value of obtaining evidence from diverse representatives of animal categories before concluding that something is generally true of a kind (e.g., Heit & Feeney, 2005; Heit, Hahn, & Feeney, 2004; Kim & Keil, 2003; Osherson et al., 1990), whereas younger children do not (Rhodes, Brickman, & Gelman, 2008; Rhodes & Liebenson, 2015). Incorporating information about category variability into representativeness judgments would be a necessary prerequisite for viewing average exemplars as representative of their kinds. Many of the features of biological reasoning discussed thus far persist across development, including tendencies towards teleological reasoning (Casler & Kelemen, 2008; Kelemen, Rottman, & Seston, 2013; Lombrozo, Kelemen, & Zaitchik, 2007; Ware & Gelman, 2014), overlooking variation across members of biological categories (Coley & Tanner, 2015; Gelman & Rhodes, 2012; Mayr, 1982; Shtulman, 2006, 2017), and thinking of characteristic animal properties in normative terms (Haward et al., 2018; Prasada & Dillingham, 2006, 2009). Nevertheless, these components of biological thought are more pronounced in early childhood. Also, as reviewed earlier, previous research has found a rather limited role for ideals in adult conceptual structure (Kim & Murphy, 2011). For these reasons, we expected to find an age-related change across childhood, from viewing ideal exemplars as most representative to viewing more average exemplars in

this manner.

The present studies first test the extent to which young children view exemplars that have extreme values of characteristic features as most representative and informative of their kinds, and how these beliefs change across development (in Studies 1 and 2). Then (in Study 3), we examine the mechanisms underlying these judgments in more detail. In all studies, we directly measured (a) which exemplars people think are most representative of natural categories and (b) which exemplars people think will provide the most generalizable information. Thus, we assessed both concept representation and concept use. In Studies 1 and 2 we also assessed representations of “the best” of a category (by asking half of participants to pick, e.g., “the best cheetah”) to directly compare these judgments to those of representativeness (by asking the other half of participants to pick the most representative cheetah; see prompts below).

## 2. Study 1

### 2.1. Participants

Children in all studies were recruited and tested in the American Museum of Natural History (AMNH) in New York City; participants were mostly from urban and suburban environments across the United States. Parents were approached within the Discovery Room at AMNH and asked if they and their child would like to participate in a brief study on how children reason about the natural world. If they agreed to participate, children were brought to a testing area within the Discovery Room. Parents completed written informed consent and children gave verbal assent. The Institutional Review Boards of the authors’ university and of AMNH approved all study procedures.

Because biological reasoning is shaped by experience and culture across development (Bailenson, Shum, Atran, Medin, & Coley, 2002; Coley, Hayes, Lawson, & Moloney, 2004; Medin et al., 2010; Ross, Medin, Coley, & Atran, 2003), we gathered information on participants’ backgrounds and experiences with nature using a parent questionnaire. In the current studies, none of these background variables significantly predicted responses or interacted with any of our key independent variables (see <https://osf.io/v5pcs/> for the questionnaire and analyses), although variation in these patterns could certainly emerge in other research if children were sampled from more diverse locations than included here.

Children were sampled from a continuous age range from 5–8 years in Studies 1 and 3 and 5–10 years in Study 2; however, we planned to consider age as a grouping variable (ages 5–6, 7–8, 9–10) because we expected different patterns across particular age groups identified in previous work (Rhodes & Liebensson, 2015). Therefore, we calculated an expected sample size of 24 per condition within each group. Grouping children into age groups also facilitated random assignment of condition within each age range and allowed us to directly compare data from children and adults. Of participants who provided location information, 212 were from urban environments, 120 from suburban environments, and 11 from rural environments. Participants in Study 1 included children ages 5–6 ( $N = 48$ ;  $M$  age = 5.89; 21 male) and 7–8 ( $N = 48$ ;  $M$  age = 7.87; 18 male) recruited from the museum. Across the studies in this research, children were 67% White, 2% Black, 13% Asian, 2% Middle Eastern or North African, and 9% more than one race; 13% were Hispanic (of any race); 6% of parents did not report demographics. Participants in Study 1 also included adults ( $N = 56$ ,  $M$  age = 33.08; 30 male) recruited from Amazon Mechanical Turk and tested via Qualtrics ([https://nyu.qualtrics.com/jfe/form/SV\\_eyecju1nHpPm0cJ](https://nyu.qualtrics.com/jfe/form/SV_eyecju1nHpPm0cJ)). Across the studies in this research, adults were 82% White, 8% Black, 8% Asian, 1% Native American or Alaskan Native, and 1% more than one race; 8% were Hispanic (of any race).

### 2.2. Methods and materials

The full study protocol for this and all studies is available at <https://osf.io/v5pcs/>; videos of the testing procedure are available for authorized users of Databrary at <https://nyu.databrary.org/volume/644>.

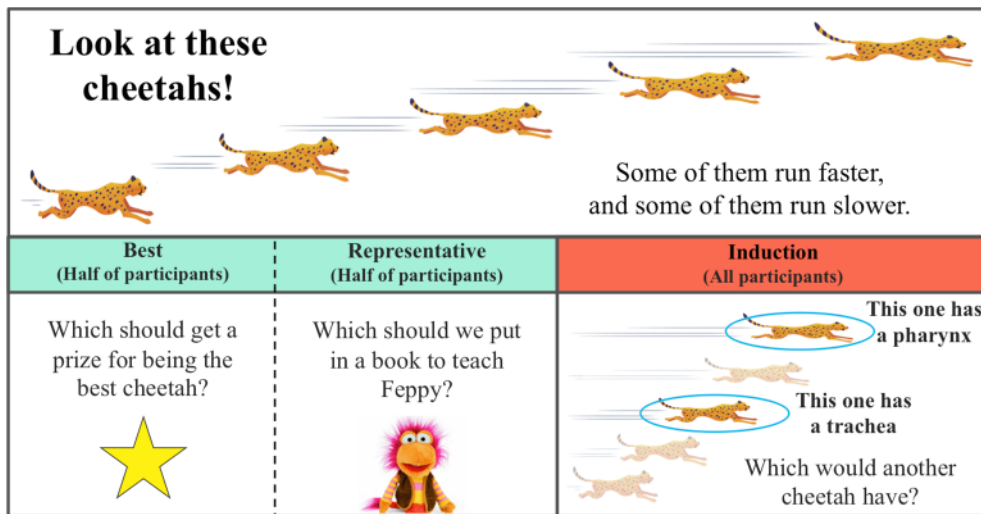
#### 2.2.1. Scale training

Children (in all studies) first completed a scale training phase to ensure they understood “most,” “least,” and “the middle” before moving on. For example, children were shown an array of five glasses that varied in how much water they contained and were asked to point to the one that had the “most,” the “least,” and the one that was in “the middle.” Children completed three items of this type to become familiar with the scales that would be used in the main portion of the experiment. Adult participants did not complete scale training as we expected all adults to understand the meaning of “most,” “least,” and “the middle.”

#### 2.2.2. Condition-specific training

After the scale training, children completed a training phase that was specific to their condition, to help them understand the experimental task. Participants within each age group were randomly assigned to either the “Best” or the “Representative” condition. In the Best condition, children practiced giving out prizes for being “the best.” For example, they were shown two ice cream cones that varied slightly in color and were told, “Look at this ice cream cone. This is the very best ice cream in the whole world. This (other) ice cream is ok but it’s not the best ice cream. Which ice cream should we give a prize to for being the best ice cream?” Children completed two items of this type to become familiar with the idea of giving prizes to the “best” member of a category.

In the Representative condition, children practiced the Representative task, which involved selecting items to put in a book to teach a novice about a category. This task has been used successfully to assess representativeness beliefs in young children previously (Rhodes et al., 2008). Children were introduced to a puppet named Feppy and were told that, “Feppy comes from another planet and



**Fig. 1.** In Study 1, all participants saw illustrations of familiar animal kinds (cheetahs, elephants, porcupines, giraffes, skunks, kangaroos, sharks, and lions), one at a time, each as a scale of five exemplars ranging from least to most extreme on a characteristic adaptive feature (e.g., faster cheetahs, stinkier skunks; top panel). We described variability in the relevant property (e.g., “This cheetah runs the fastest, this cheetah runs the slowest, and this cheetah is in the middle, it runs faster than some cheetahs and slower than some other cheetahs.”) In the testing phase, participants in the Best condition gave a prize to the “best” exemplar (lower-left panel); participants in the Representative condition chose an exemplar to put in the book to teach a novice (a puppet named “Feppy” for children, “the stranger” for adults) about the category (lower-center panel). Participants in both conditions then completed a forced-choice induction question about each animal kind (lower-right panel) in which they heard that the most extreme exemplar had one unfamiliar property while the average exemplar had a different one; they chose which of the two properties another member of the kind would have. Study 2 used the same method as Study 1 except that a different prompt was used for the Representative condition, with participants asked to select the “real, cheetah-y cheetah.”

doesn't know lots of the things that you know. But he really wants to learn about our world! Can you help me teach Feppy about our world? Let's imagine that you're going to make a book about some different things so Feppy can learn.” Then, for the first practice item: “Let's teach Feppy about triangles. This shape has three sides and this shape has four sides. Which do you want to put in the book to teach Feppy about triangles?” This warm-up was only intended for children to understand the idea of selecting examples to put in a teaching book, the actual test items followed a different structure, as described below.

Adult participants did not complete condition-specific training as we expected all adults to understand the concepts of giving out prizes or selecting examples to put in a book to teach someone. In the Best condition, adults were told, “Imagine you have a bunch of prizes to give out to things for being the best. You are giving out prizes to things for being the very best.” In the Representative condition, adults were told, “Imagine that you met a stranger from another planet who wanted to learn about our world. In order to teach the stranger, you are making a book about some different animals.” The structure of test items was identical for child and adult participants.

### 2.2.3. Test items

For the test phase in both conditions, participants of all ages made judgments about eight familiar animal categories (see Fig. 1). Each category was presented as a scale of five exemplars varying in a key, characteristic property of the category (e.g., five giraffes varying in neck length; five cheetahs varying in speed). Half of these extreme values were not desirable from the human perspective (e.g., “the stinkiest skunk”, “the shark with the sharpest teeth”), allowing us to examine ideals in the sense of maximal values on characteristic adaptive properties, rather than those that better meet human goals.

In the Best condition, participants selected which exemplar should get a prize (a gold star) for being the “best” of its category (e.g., “the best cheetah”). We expected participants of all ages to choose exemplars that were extreme on the relevant property to receive prizes (i.e., “the best cheetah” would be the very fastest one), though we suspected these judgments might vary by valence, with some selecting the *least* stinky skunk as “best” rather than the *most* stinky skunk (Kim & Murphy, 2011).

In the Representative condition, participants were asked to select the exemplar that most clearly represents the category by choosing which exemplar to put in a book to teach a novice (“Feppy” for children, “the stranger” for adults) about the category. If ideal exemplars are more central to early representations of biological categories, younger children should select examples that are extreme on characteristic properties as most representative, and such selections should decline with age (in favor of more average selections).

For each item, after participants completed the condition-specific task, they were asked an induction question (identical across both conditions) before proceeding to the next animal trial. Participants were asked to predict whether a novel property found in the most extreme exemplar (e.g., the fastest cheetah) or the average exemplar (e.g., the average speed cheetah) would generalize to another, not-pictured exemplar. People often view representative exemplars as providing the most generalizable information



(Osherson et al., 1990; but see Rein et al., 2010 for an alternate interpretation), so our predictions for inductive inferences were the same as for representativeness judgments: Younger children should prefer to generalize from extreme exemplars, with a shift toward average exemplars more with age. On all measures, children indicated their answers by pointing at one exemplar on each scale. Adults clicked a button onscreen corresponding to one exemplar on each scale.

## 2.2.4. Coding and analysis

All data collection sessions with children were videotaped for reliability—children's responses (which exemplar they pointed to on the 5-point scale) were coded "live" by the experimenter administering the task, and then coded separately from video by a different researcher. Inter-rater agreement was 98%, with disagreements resolved by the first author reviewing the video file.

We analyzed participants' selections on the main task (assessing their beliefs about the Best or most Representative exemplars, by condition) using the lme4 package in R version 3.4.1. We included random effects for participants and items in all models to control for the variance associated with these factors without data aggregation (Judd, Westfall, & Kenny, 2012). We implemented Linear Mixed Models (LMM), testing for the main and interactive effects of age group, condition, and item valence (whether *more* of the characteristic property was objectionable or not from the human perspective). We report the F tests from the LMER results using the lmerTest package (Type III Wald F tests with Kenward-Roger degrees of freedom approximation). Means are reported as the average exemplar selected on the 1–5 scales with 95% confidence intervals. Higher numbers indicate more extreme exemplars on the characteristic property.

Induction decisions were a forced choice between the average exemplar (coded as 0) and the most extreme (coded as 1). We analyzed these data using a Generalized Linear Mixed Model (GLMM) to specify a binomial distribution, with the main and interactive effects of age group, condition, and item valence as predictors. We report Wald chi-square tests from the GLMER results; means are reported as the probability of selecting the extreme exemplar with 95% confidence intervals. All data and analysis code are available at <https://osf.io/v5p5cs/>.

## 2.3. Results and discussion

Participants of all ages selected extreme exemplars as the "best" of their categories (e.g., they gave a prize to the fastest cheetah), but judgments of which exemplars were most representative changed across age (see Fig. 2). Younger children selected extreme exemplars as most representative, and these selections declined with age in favor of more average exemplars. The age  $\times$  condition interaction was reliable ( $F(2, 146) = 5.33, p = 0.006$ ) as were subsumed main effects of age ( $F(2, 146) = 21.29, p < 0.001$ ) and condition ( $F(1, 146) = 25.95, p < 0.001$ ). Younger children were equally likely to pick extreme exemplars when asked to select the best as when asked to pick the most representative ( $F(1, 46) = 0.10, p = 0.75$ ), whereas children ages 7–8 ( $F(1, 46) = 17.40, p < 0.001$ ) and adults ( $F(1, 54) = 20.31, p < 0.001$ ) selected more extreme exemplars when asked to select the best than when

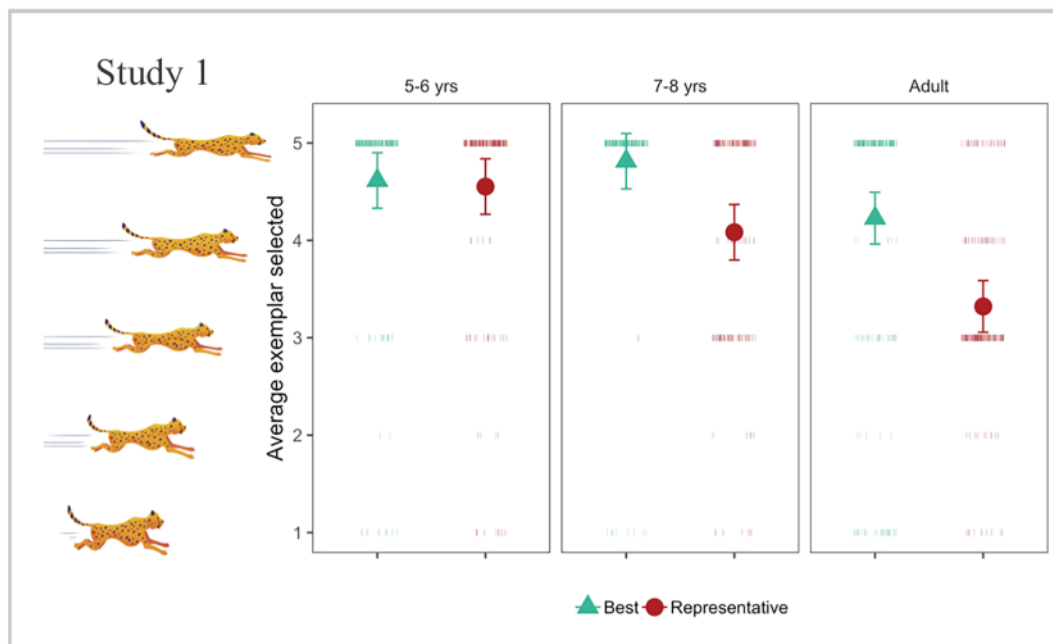


Fig. 2. Average exemplar selected (as the best or most representative) in Study 1 ( $N = 152$ ), by age group and condition. Higher values represent exemplars that are more extreme on a characteristic adaptive property (e.g., faster cheetahs, stinkier skunks) across all eight animal trials. Exemplar 3 represents selecting the animal in the middle of the scale. Large shapes represent group means by age group and condition; error bars show 95% Confidence Intervals. Small lines are responses to individual trials.

asked to pick the most representative. Tukey post-hoc tests indicated that representativeness choices differed significantly ( $p < 0.05$ ) between each age group.

Younger children also viewed extreme exemplars as providing more generalizable information. They were more likely to say, for example, that an unseen cheetah would be like the fastest cheetah (coded as 1) than the average-speeded cheetah (coded as 0;  $M = 0.65$ , 95% CI [0.51, 0.77]). This tendency declined with age: Older children and adults more often generalized from average (instead of extreme) exemplars (ages 7–8:  $M = 0.41$ , 95% CI [0.28, 0.56]; adults:  $M = 0.10$ , 95% CI [0.06, 0.18], Wald  $X^2(2) = 22.68$ ,  $p < 0.001$ ). Tukey post-hoc tests indicated that informativeness choices differed marginally between the two groups of children ( $p = 0.05$ ) and differed significantly between 7–8 year olds and adults ( $p < 0.001$ ). The youngest children chose extreme exemplars on induction questions significantly more often than predicted by chance ( $p = 0.03$ ), 7–8 year olds were at chance ( $p = 0.25$ ), and adults were significantly below chance ( $p < 0.001$ ).

Considering the valence of the properties (e.g., whether an extreme value was desirable or not from a human perspective), participants were slightly less likely to choose extreme values as “the best” when the properties were objectionable from a human perspective ( $F(1, 6.02) = 7.51$ ,  $p = 0.03$ ; negative valence, e.g., the stinkiest skunk,  $M = 4.34$ , 95% CI [4.09, 4.60]; positive valence, e.g., the fastest cheetah,  $M = 4.76$ , 95% CI [4.51, 5.02]). The effect of valence on judgments about which exemplars were the best did not interact with age,  $F(2, 523) = 1.2$ ,  $p = 0.3$ ). Critically, there were no effects of valence on representativeness choices ( $F(1, 6.01) = 0.31$ ,  $p = 0.60$ ) or induction (Wald  $X^2(1) = 1.32$ ,  $p = 0.25$ ).

Overall, Study 1 revealed that younger children view exemplars with more extreme values on characteristic adaptive properties as representative and informative. Older children and adults, in contrast, view more average exemplars in this manner.

### 3. Study 2

The purpose of Study 2 was to replicate the pattern found in Study 1 with a new sample of participants, using a different prompt for eliciting representativeness beliefs. The prompt used in Study 1 asked participants to select exemplars to put in a book to teach someone about the categories; thus, one possibility is that participants selected the exemplars they viewed as having the most pedagogical value instead of ones they viewed as most representative more generally (Csibra & Gergely, 2009; Rhodes, Bonawitz, Shafto, Chen, & Caglar, 2015; Shafto, Goodman, & Griffiths, 2014; Xu & Tenenbaum, 2007). Although these judgments might coincide, it is also possible that children might think the fastest cheetah would be helpful for teaching about cheetahs, for example (in order to help a learner differentiate it from other similar animals), but still view an average-speeded one as more representative. To address representative beliefs more directly, we built on classic studies in the adult literature on representativeness (Rosch & Mervis, 1975); participants in the Representative condition of Study 2 were asked to select, e.g., the “real cheetah-y cheetah.”

#### 3.1. Participants

As in Study 1, participants in Study 2 included children ages 5–6 ( $N = 52$ ;  $M$  age = 5.98; 25 male) and ages 7–8 ( $N = 54$ ;  $M$  age = 7.96; 26 male) and adults (57,  $M$  age = 33.24; 39 male). Given that the oldest children tested in Study 1 (7–8 year olds) showed a pattern of responses that differed from that of adults, in Study 2 we also included a sample of 9–10 year old children ( $N = 48$ ;  $M$  age = 10.1; 24 male) to more fully track the pattern of developmental change across childhood. All participants were recruited and tested as in Study 1.

#### 3.2. Method

As in Study 1, participants were randomly assigned to the Best or Representative condition. Children completed the same scale training as in Study 1. Children in the Best condition also completed the same task-specific training, whereas the task-specific training in the Representative condition was updated for the new task in this condition. Similar to classic studies in the adult concepts literature (Kim & Murphy, 2011; Rosch & Mervis, 1975), children were introduced to the idea of a “real red-y red.” They then practiced selecting the “real lamp-y lamp” from a choice of a more familiar lamp and a more unusual-looking lamp, and the “real chair-y chair” from a similar set of options (for full protocol, see <https://osf.io/v5pcs/>). Adults in the Representative condition completed the same task-specific training and practice items as children to ensure they understood the task.

After the trainings, participants completed their condition-specific task (selecting the best or most representative exemplar) for a series of eight familiar animals, as in Study 1. Also, after each condition-specific task, participants completed the forced-choice inductive inference task before moving on to the next animal trial as in Study 1. Data were processed and analyzed as for Study 1.

#### 3.3. Results and discussion

The pattern of results was identical to Study 1 (see Fig. 3). Participants of all ages selected extreme exemplars as “the best” of their categories, but judgments of which exemplars were most representative changed across age: Younger children selected extreme exemplars as most representative, and such selections declined with age in favor of more average exemplars. The age  $\times$  condition interaction was reliable ( $F(3, 203) = 9.40$ ,  $p < 0.001$ ) as were subsumed main effects of age ( $F(3, 203) = 6.8$ ,  $p < 0.001$ ) and condition ( $F(1, 203) = 44.69$ ,  $p < 0.001$ ). Younger children were just as likely to pick extreme exemplars when asked to select the best exemplar as when asked to pick the most representative exemplar ( $F(1, 50) = 0.42$ ,  $p = 0.52$ ), whereas children ages 7–8 ( $F(1, 52) = 7.88$ ,  $p = 0.007$ ), ages 9–10 ( $F(1, 46) = 33.28$ ,  $p < 0.001$ ), and adults ( $F(1, 55) = 42.64$ ,  $p < 0.001$ ) selected more extreme

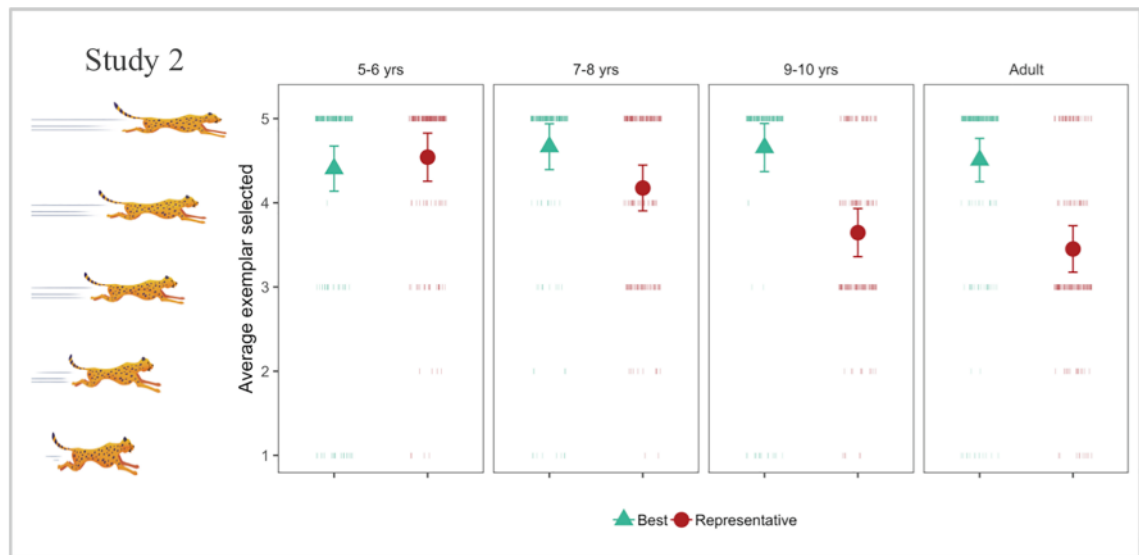


Fig. 3. Average exemplar selected (as the best or most representative) in Study 2 ( $N = 211$ ), by age group and condition. Higher values represent exemplars that are more extreme on a characteristic adaptive property (e.g., faster cheetahs, stinkier skunks) across all eight animal trials. Exemplar 3 represents selecting the animal in the middle of the scale. Large shapes represent group means by age group and condition; error bars show 95% Confidence Intervals. Small lines are responses to individual trials.

exemplars when asked to select the best than when asked to pick the most representative. Tukey post-hoc tests indicated that representativeness choices differed significantly ( $p < 0.05$ ) between all three groups of children, however 9–10 year olds' choices did not differ significantly from those of adults ( $p = 0.67$ ).

Also as in Study 1, induction items revealed that generalizations based on the ideal (instead of average) exemplar declined with age (Wald  $X^2(3) = 13.84$ ,  $p = 0.003$ ; ages 5–6:  $M = 0.52$ , 95% CI [0.38, 0.67]; ages 7–8:  $M = 0.43$ , 95% CI [0.30, 0.58]; ages 9–10:  $M = 0.22$ , 95% CI [0.13, 0.35]; adults:  $M = 0.17$ , 95% CI [0.10, 0.27]). Informativeness choices did not differ between the two younger age groups ( $p = 0.83$ ) or between 9–10 year olds and adults ( $p = 0.89$ ), but differed marginally between 7–8 year olds and 9–10 year olds ( $p = 0.08$ ). The induction choices of 5–6 year olds and 7–8 year olds did not differ significantly from chance ( $p > 0.05$ ), whereas 9–10 year olds and adults chose extreme exemplars on induction items significantly less often than predicted by chance ( $p < 0.001$ ).

With regard to valence, participants' choices for the best exemplars were again less extreme when the property was negative from a human perspective ( $F(1, 6.01) = 6.81$ ,  $p = 0.04$ ; negative valence,  $M = 4.17$ , 95% CI [3.98, 4.36]; positive valence,  $M = 4.37$ , 95% CI [4.18, 4.55]), with no valence effects on representativeness ( $F(1, 6.01) = 0.03$ ,  $p = 0.86$ ) or induction (Wald  $X^2(1) = 1.11$ ,  $p = 0.29$ ).

The data from Study 2 thus confirm the pattern of age-related change found in Study 1. Younger children viewed exemplars with extreme values on characteristic adaptive properties as the most representative and informative exemplars of their categories, with a shift towards more average exemplars with age. The fact that we found the same pattern of age-related change across two very different tasks for assessing representativeness suggests that the focus on idealized exemplars is a robust component of early conceptual structure. We also found similar patterns on our measures of both concept representation and concept use (with the induction task). For concept use, in both studies 1 and 2, we found consistent age-related changes towards viewing average exemplars as providing more generalizable information. In both studies, adults reliably chose to generalize from average over ideal exemplars, and in Study 2, we also found this pattern among the oldest children. One thing to note, however, is that while generalizations from ideal exemplars were more common among younger children than adults in both studies, in Study 1 younger children reliably generalized from ideal over average exemplars, whereas in this Study, they did not show a reliable pattern across trials.

#### 4. Study 3

Study 3 examined the mechanisms underlying participants' choices. We also addressed low-level alternative accounts for why younger children responded as they did in Studies 1 and 2. For example, perhaps children had a simple bias to select one side of the scale or to select exemplars with the most of a salient perceptual property. To address these issues in Study 3, we introduced participants to novel, realistic (but fictional) animal categories. Each new animal category was shown as a scale of exemplars varying along one easily perceptible feature as in Studies 1 and 2. Unlike Studies 1 and 2, however, because the categories and properties were novel, participants should not (at baseline) view the variable feature as normatively tied to category membership. Thus, if younger children in Studies 1 and 2 selected extreme exemplars because of their normative belief about category structure, they should not do so here. If they did so only because they prefer exemplars that show *more* of something or that are on one side of the



scale, on the other hand, their responses should be similar to Studies 1 and 2 at baseline.

We also evaluated the contributions of two theoretically relevant mechanisms that could underlie the belief that extreme exemplars are most representative: (a) that such exemplars best fulfill what category members are *supposed* to be like (e.g., in the case of biological kinds because they increase adaptive fit) or (b) because extreme exemplars are more helpful for differentiating categories from one another. As described in the Introduction, this research was motivated by a hypothesis in line with mechanism (a); thus far, however, while we have found evidence that young children view extreme exemplars as representative and informative, we have not yet tested if they do so *because* they view these exemplars as best fulfilling category ideals. Directly testing this mechanism was one goal of Study 3. To consider the contribution of mechanism (a), in some conditions (but not others), participants learned that the target feature had important, functional consequences for a characteristic, species-specific adaptive behavior (e.g., a longer snout allowed the animal to get food more successfully in a specific way). We expected that receiving this functional information would lead participants to reason about variation within the target properties in terms of increasing adaptive fit of the organism (Kelemen, 1999, 2003; Kelemen et al., 2005; Springer & Keil, 1989). This functional information about what helps category members enhance survival would highlight the feature as an important aspect of how category members are supposed to be (Haward et al., 2018; Prasada & Dillingham, 2006, 2009) in a way that would not be obvious based on perceptual variation alone for these novel animals, and would define a salient ideal for each novel animal category (e.g., the animal with the very longest snout). If younger children in Studies 1 and 2 viewed extreme exemplars as representative *because* they view them as more ideal category members from this perspective, they should do so in Study 3 when they receive this functional information, but not otherwise. To further address possible low-level mechanisms, we also varied the nature of the novel properties so that having *more* of a target property facilitated the key behavior (e.g., longer snouts helped dig for food) for half of the items, whereas having *less* did so for the other half (e.g., smaller ears were better for hiding from predators).

In designing Study 3, we also considered an alternate (or additional) reason why young children could value extreme exemplars (mechanism b)—that such exemplars are particularly helpful for differentiating categories from one another and thus are useful at earlier stages of concept acquisition (Ameel & Storms, 2006; Davis & Love, 2010; Goldstone et al., 2003; Kim & Murphy, 2011; Levering & Kurtz, 2006). From this perspective, children might view the fastest cheetah as a particularly good representative of cheetahs, for example, not because they think that this cheetah best fulfills a category ideal, but because such a cheetah would be easier to differentiate from a lion.

To consider the contribution of this mechanism, we varied whether participants were introduced to a relevant (also novel) contrast category in addition to the target category, or not. We predicted that if participants view extreme exemplars as representative because they are helpful for differentiation, they should choose exemplars that are more distinct from contrast categories differing in the critical dimension when such contrasts are present (Ameel & Storms, 2006; Davis & Love, 2010; Kim & Murphy, 2011; Levering & Kurtz, 2006; Rosch & Mervis, 1975). For instance, if neck length is helpful because it can differentiate giraffes from zebras, then people should view longer-necked giraffes as particularly representative and informative when zebras are present. From this perspective, in a world where no one knew anything about giraffes in advance, people might choose longer-necked giraffes as representative when introduced to the category while also learning about zebras, but they would choose shorter-necked giraffes when introduced to them along with another animal whose necks were even longer. To examine the relative contributions of mechanisms (a) and (b), we crossed these factors in a  $2 \times 2$  experimental design (Fig. 4). The condition in which children received neither functional information nor a contrast category served as the baseline control condition.

In order to test how the mechanisms underlying participants' responses might change across age, we again included children, ages 5–6 years and ages 7–8 years, and adults. As 9–10 year old children's responses did not differ from adults' on either measure in Study







These are tulvers!					
Look, some have longer snouts and some have shorter snouts.					
					
	1	2	3	4	5
Functional (Half of participants)	No functional (Half of participants)	Contrast (Half of participants)	No contrast (Half of participants)	Representative (All participants)	Sample selection (All participants)
Tulvers eat bugs that live in deep holes underground, so longer snouts are better for getting more food to eat	Tulvers live in the desert where there's lots of soft sand, so they all have nice soft places to sleep	This other kind of animal lives nearby 	All day long they don't see any other kinds of animals	Which is the <i>real</i> , tulver-y tulver?	Which should we look at to find out if tulvers have a pharynx?

Fig. 4. In Study 3, after completing a scale-training phase as in Studies 1 and 2, participants made representativeness and informativeness judgments about eight novel animals, one at a time, each presented as a scale of five exemplars ranging from least to most extreme on a single feature (e.g., snout length). Depending on condition in a  $2 \times 2$  design, participants heard different types of information about each animal kind: functional information (given, not given), and contrast category information (present, not present). All participants then selected which exemplar was most representative (e.g., “the real, tulver-y tulver”) and which (of all five exemplars) was most informative about the category (e.g., “which should we look at to learn about tulvers?”). Responses were scored from 1 to 5, with higher numbers indicating exemplars that better met functional goals (when functional information was present) and lower numbers indicating exemplars that were more contrastive (when a contrast category was present).



2, we did not include this age group in Study 3. Based on Studies 1 and 2, one possibility is that ideals are important only in early conceptual structure of natural kinds, and not at all in the natural kind concepts of older children or adults. Alternately, perhaps a propensity to reason about biological variation in terms of teleological ideals persists in conceptual structure across development but plays a more limited role (see, for example, Casler & Kelemen, 2008; Kelemen et al., 2013; Ware & Gelman, 2014). In this case, we might see its consequences more when adults are learning about new categories (Davis & Love, 2010; Levering & Kurtz, 2006) or when the importance of properties normatively tied to biological categories is particularly salient (Barsalou, 1985; Burnett et al., 2005; Lynch et al., 2000; Haward et al., 2018; Prasada & Dillingham, 2006, 2009).

#### 4.1. Participants

Participants included children aged 5–6 ( $N = 96$ ;  $M$  age = 5.92; 51 male) and 7–8 ( $N = 100$ ;  $M$  age = 7.98; 42 male), recruited as in Studies 1 and 2. We also included adults ( $N = 100$ ,  $M$  age = 35.43; 64 male), recruited with Amazon Mechanical Turk and tested using the Testable.org online testing platform ([testable.org/t/781f4c9a0](https://testable.org/t/781f4c9a0)).

#### 4.2. Methods and materials

Materials for Study 3 were color illustrations of eight novel animals, each presented as a scale of five exemplars ranging from least to most extreme (or vice versa) along one property dimension (e.g., snout length). Varying properties were neutral in valence so that participants would not have strong prior beliefs about which features might help fulfill category goals (e.g., eye size was not used because larger eyes are considered more ideal; Costa & Corazza, 2006). Each animal kind was given a novel name (from the NOUN database; Horst & Hout, 2016). Novel animals were designed to look realistic but not too similar to any familiar animal kind. If children commented that a novel animal resembled a familiar animal during testing, they were told, “Hmm, this is a *different* kind of animal called [name].”

Children in all conditions first completed the scale training as in Studies 1 and 2, followed by the condition-specific training for the Representative condition from Study 2; adult participants completed only the representativeness training. All participants were then introduced to the series of eight fictional animals in a manner that was specified by their condition. Participants within each age group were randomly assigned to one of four conditions, following a 2 (Functional ideals information: given, not given)  $\times$  2 (Contrast category: present, not present) factorial design (see Fig. 4 for examples of how the categories were introduced in each condition, and <https://osf.io/v5pcs/> for the full study protocol). The order of Functional ideals and Contrast category information did not vary across conditions.

After they were introduced to the novel categories, as specified by their assigned condition, all participants were asked to select the most representative exemplar using the same method as Study 2 (e.g., participants selected the “real tulver-y tulver”). Then, on each trial, to assess which exemplar they thought was the most informative, participants were asked to select which, of all five exemplars on the scale, they would choose to look at to learn something new about the category (e.g., “which should we look at to learn about tulvers?”). Participants completed the representativeness question and informativeness question for each novel animal item before moving onto the next item.

Representativeness judgments were analyzed as in Studies 1 and 2. The format of the informativeness question differed from Studies 1 and 2 (i.e., participants chose which they would like look at to learn from out of all five possible exemplars, rather than making a forced choice between the most extreme or average exemplar as in Studies 1 and 2); we made this change to ensure that our measures did not lead participants to focus on one end of the scale in the absence of functional information. These data were therefore analyzed in the same way as the representativeness judgments, instead of via binomial models as in Studies 1 and 2. Animal trials in which the most functionally valuable exemplar was on the left of the scale rather than the right (half of all trials) were reverse-coded so that, for all trials, higher numbers indicated more functionally valuable exemplars and lower numbers indicated more contrastive exemplars.

#### 4.3. Results and discussion

In the baseline condition, when participants were introduced to the five exemplars of the novel categories that varied perceptually (e.g., “tulvers” varying in snout length) but were not told about functional consequences or contrastive value of the property, participants of each age selected exemplars around the middle of the scale (i.e., 3) as both most representative (5–6 year olds:  $M = 3.13$ , 95% CI [2.82, 3.43]; 7–8 year olds:  $M = 3.12$ , 95% CI [2.8, 3.44]; adults:  $M = 3.08$ , 95% CI [2.77, 3.39]) and informative (5–6 year olds:  $M = 2.9$ , 95% CI [2.61, 3.19]; 7–8 year olds:  $M = 3.31$ , 95% CI [2.99, 3.62]; adults:  $M = 3.05$ , 95% CI [2.75, 3.35]). Although participants of all ages picked more extreme values when the value of the feature increased in a least-to-most direction (and this tendency was particularly pronounced among the youngest children), this variable did not interact with functional information and the effects of functional information (described below) remained unchanged even when direction information was incorporated into the statistical models (see <https://osf.io/v5pcs/> for details). These patterns indicate that children’s responses in Studies 1 and 2 cannot be explained by a low-level bias to always pick exemplars on one side of the scale, or to always simply select exemplars that had “more” of something, paving the way to consider the role of more theoretically meaningful mechanisms.

Consistent with the possibility that participants judge extreme exemplars as representative when they display extreme values of properties that are normatively important for category membership (in this case, because they increase adaptive fit), participants who received functional information about the properties chose more extreme (i.e., functionally ideal) exemplars as representative of their

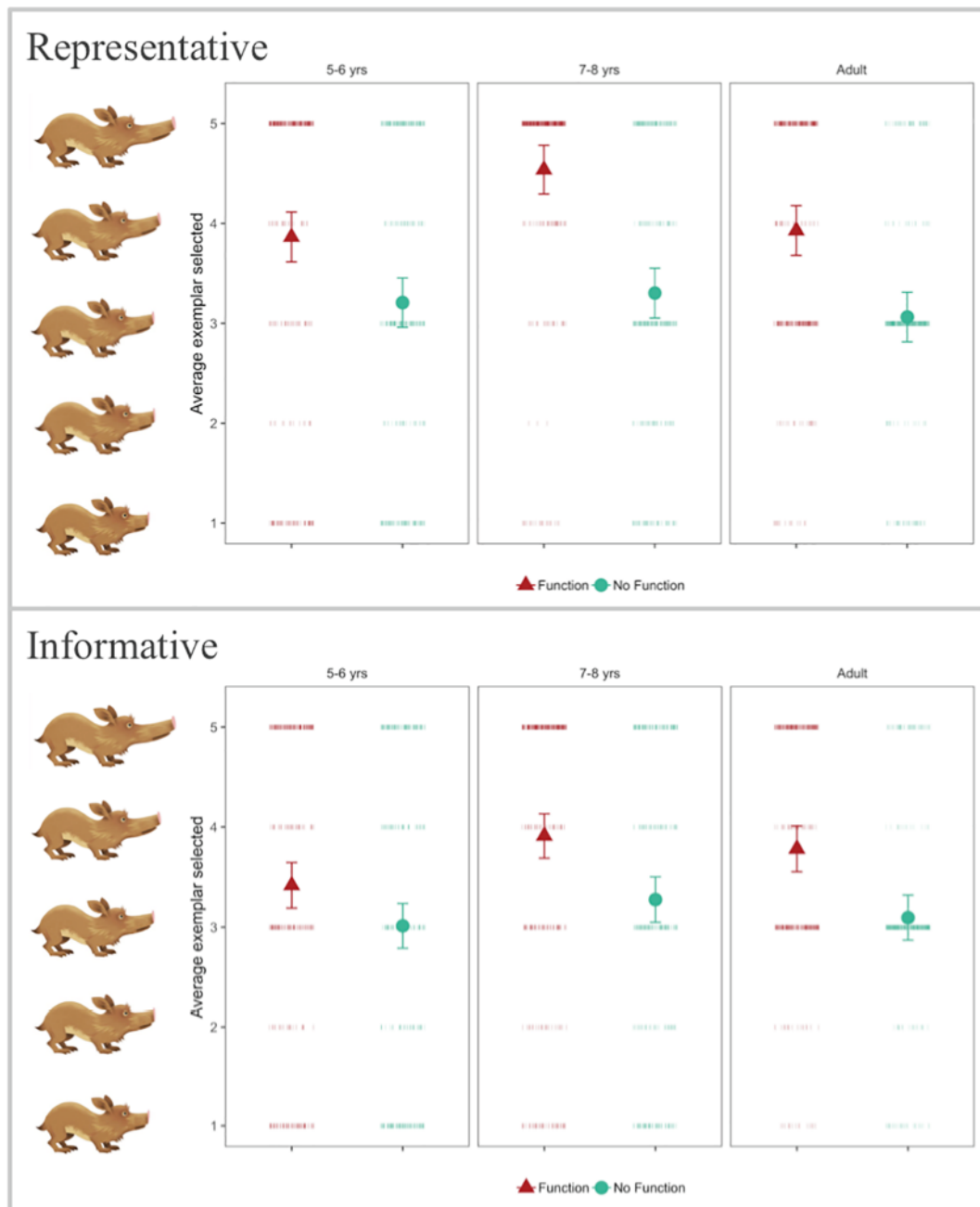


Fig. 5. Average exemplar selected, by age group and functional ideal information, collapsed across both levels of contrast category information ( $N = 296$ ). Higher values represent more functionally valuable exemplars. Exemplar 3 represents selecting the animal in the middle of the scale. Large shapes are means by age group and condition across all eight trials; error bars show 95% Confidence Intervals. Small lines are responses to individual trials. The top graph shows representativeness responses (e.g., “which is the *real*, *tulver-y tulver?*”); the bottom graph shows informativeness responses (e.g., “which would you look at to learn about tulvers?”).

kinds than those who did not receive this information (and instead had access only to perceptual variation; main effect of functional ideals information,  $F(1, 284) = 142.64$ ,  $p < 0.001$ ). Although the effect of functional information varied by age group ( $F(2, 284) = 4.7$ ,  $p = 0.01$ ), at each age participants picked more extreme values when they were given functional information than when they were not (Fig. 5). The interaction reflected that this effect was more pronounced among children ages 7–8 than either other group, but the effect of functional information was similar and reliable at each age considered separately (ages 5–6:  $F(1, 92) = 20.99$ ,  $p < 0.001$ ; ages 7–8:  $F(1, 96) = 110.17$ ,  $p < 0.001$ ; adults:  $F(1, 96) = 39.57$ ,  $p < 0.001$ ).

Participants were also more likely to select extreme exemplars on adaptive properties as most informative (revealed by their

decisions of which exemplars to examine to learn about the category) when they received functional ideals information about the properties than otherwise (main effect of functional information,  $F(1, 284) = 48.07, p < 0.001$ ). Although children ages 7–8 selected more extreme exemplars overall ( $M = 3.60$ , 95% CI [3.43, 3.78]) than 5–6 year olds ( $M = 3.22$ , 95% CI [3.04, 3.40]) or adults ( $M = 3.45$ , 95% CI [3.27, 3.62]), the effect of functional information on informativeness judgments did not vary by age group ( $F(2, 284) = 1.06, p = 0.35$ ) and was reliable at each age examined separately (ages 5–6:  $F(1, 92) = 9.36, p = 0.003$ ; ages 7–8:  $F(1, 96) = 15.59, p < 0.001$ ; adults:  $F(1, 96) = 25.90, p < 0.001$ ).

Inconsistent with the possibility that people view extreme exemplars as representative because they are more helpful for differentiating categories from one another, participants were not more likely to select extreme exemplars that would be helpful for differentiation in the presence of contrast categories (in fact, they were slightly less likely to do so;  $F(1, 284) = 4.06, p = 0.045$ ; contrast:  $M = 3.74$ , 95% CI [3.54, 3.94]; no contrast:  $M = 3.59$ , 95% CI [3.39, 3.78]; note that items were scored such that lower numbers indicated greater contrastive value). Contrast category information did not interact with information about functional ideals ( $F(1, 284) = 0.05, p = 0.82$ ; see <https://osf.io/v5pcs/> for additional analyses of this factor). There were no effects of the presence of contrast categories on choices on which exemplar was most informative.

Overall, the data from Study 3 indicate that children view extreme exemplars as most representative and informative because they best match their views of what category members are supposed to be like, by supporting inferences about adaptive fit. Further, we found that in some circumstances the tendency to treat extreme exemplars in this manner persists into adulthood. In this study, we found similar patterns for both concept representation and concept use—in both cases, children and adults selected more extreme exemplars as representative and informative if they thought of the extreme exemplars as better meeting adaptive goals. Also, in the presence of functional information only, children and adults selected exemplars as representative and informative that were reliably more extreme than the midpoint of the scale (see Fig. 5).

## 5. General discussion

Category members viewed as highly representative of their kinds (e.g., robin for the category of birds) take on powerful roles in learning, memory, and reasoning. Thus, it is critical to determine how people judge which category members are most representative. The current studies revealed developmental change in this important component of category structure for natural kinds and provided evidence of an early-emerging and persistent role for functional, adaptive ideals in biological category representations. In Studies 1 and 2, younger children chose exemplars with more extreme values of characteristic normative properties as representative and informative, whereas adults chose average exemplars. The shift from more ideal to more average exemplars occurred gradually between ages 5–10. In Study 3, however, participants of all ages viewed relatively extreme exemplars as most representative and informative of novel animal kinds after learning that the relevant properties were functionally connected to the adaptive fitness of the animal. Overall, these findings suggest that people sometimes think of biological categories in terms of ideal exemplars, with the extent to which they do so varying across development and category knowledge. Thus, this research reveals an early-emerging and persistent feature of conceptual representation for biological kinds—a tendency to ignore within-category variation and focus instead on narrow views of how category members *should* be.

The patterns of developmental change documented here could underlie a range of age-related findings that have never before been explained in a coherent framework, including discontinuities in inductive reasoning (Rhodes & Liebenson, 2015), sample selection (Rhodes et al., 2008), and the ability to understand adaptation by natural selection (Kelemen, 1999). For example, adults often view diverse samples of evidence (e.g., a robin, eagle, and penguin) as providing a stronger basis of generalization to broader categories (e.g., to all birds) than less diverse samples (e.g., three robins; Heit & Feeney, 2005; Heit et al., 2004; Kim & Keil, 2003; Osherson et al., 1990). Although young children are capable of tracking and reasoning about sample diversity (Heit & Hahn, 2001; Rhodes & Liebenson, 2015), they often treat diverse and non-diverse samples of evidence as equivalently informative when reasoning about familiar natural kinds until around age 9 (Gutheil & Gelman, 1997; Lopez, Gelman, Gutheil, & Smith, 1992; Rhodes et al., 2008).

The present results provide a compelling framework for understanding these findings—if young children are focused on what members of these categories *should* be like, perhaps they seek out samples that best match their idealized prototypes to learn from instead of samples that provide coverage of within-category variation. In this way, age-related discontinuities in inductive reasoning do not reflect fundamental changes in cognitive skills or computational abilities, but occur because the changing nature of children's representations of biological categories leads them to value different types of evidence in the world. Consistent with this view, even when reasoning about more variable examples of natural categories than in the present research, young children prefer to seek out highly prototypical samples (e.g., two golden retrievers) over highly diverse ones (e.g., a golden retriever and a Chihuahua)—a tendency that reverses across age (Rhodes et al., 2008; Zhong, Lee, Huang, & Mo, 2014). Also, the extent to which young children view particular prototypical examples as standing in for the category as a whole (e.g., the strength of children's beliefs that robins represent the whole category of birds) is negatively correlated with their tendency to view diverse samples as more informative than non-diverse ones (Rhodes & Liebenson, 2015). Because these previous studies used stimuli that varied along multiple property dimensions, however, it was not possible to conclusively determine that children preferred highly prototypical samples *because* they were more ideal. Thus, the current studies examined this question directly by limiting the variation in the stimulus sets to a single dimension. The specific manner in which idealized prototypes (of the sort documented in the present work) shape how children use categories to learn about the world will need to be explored more fully in future work.

The current data raise several important questions about the nature of conceptual development: Why are ideals particularly central in young children's natural kind representations? Why do they become less so across childhood for familiar categories? In the



present studies, young children viewed exemplars with extreme feature values as the clearest representatives of both familiar and novel animal categories, whereas older children and adults did so only for novel categories. One interpretation of the data regarding familiar animal kinds is that extreme feature values are particularly helpful at earlier stages of category learning (Davis & Love, 2010; Levering & Kurtz, 2006; Smith & Minda, 1998). For example, the fastest cheetah might be easiest to differentiate from another, similar animal, and relying on running speed to distinguish cheetahs could be a more important strategy when first learning about them (as in early childhood). From this perspective, developmental changes documented here for familiar animals could reflect age-related changes in category expertise.

Although increases in category expertise could explain why extreme exemplars became less salient across age in Studies 1 and 2, contrastive learning mechanisms alone cannot account for the full pattern of findings in Study 3. When participants were introduced to novel categories, they did not view exemplars with more extreme perceptual features as representative or informative at baseline, or even when those features were useful for differentiating between similar categories. Instead, participants did so only when they also had information that the properties helped fulfill special category-specific adaptive needs (e.g., finding food in a species-specific manner). Thus, participants viewed exemplars with extreme feature values as better representatives only once they viewed the relevant features as part of how category members *should* be within a relevant causal framework (in this case, the teleological goals of biological kinds).

On this account, the age-related changes documented here and the mixed findings in the literature could reflect developmental and contextual variability in the salience of ideals relative to other types of information. Specifically, information about within-species variation may be less salient to young children, due both to a heightened emphasis on prescriptive norms and to limitations in their knowledge of biological category variation. For example, to a child, the idea that “cheetahs run fast” might mean that all proper cheetahs should have the same exceptional speed, so the most representative cheetah would be the one that best illustrates this capacity (Cimpian, Brandone, & Gelman, 2010; Gelman, Leslie, Was, & Koch, 2015). Although adults could share young children’s intuition that “cheetahs run fast,” and that they *ought* to run fast, adults might balance this intuition against knowledge that actual cheetahs vary in their abilities (Shtulman & Harrington, 2016). How adults balance these opposing considerations could also vary across contexts depending on their salience. Thus, in Studies 1 and 2 adults relied more heavily on their understanding of variation, but in Study 3, when they did not have prior experience with the categories and a category ideal was made especially salient within the experiment itself, they relied more on normative beliefs. On this account, young children rely on category ideals across a broader range of contexts both because they have less experience with biological variation and because separating out prescriptive and descriptive category information might generally be more challenging for them.

The various explanations for why young children might focus on functional ideals of biological kinds suggest different possibilities regarding the extent to which the age-related trajectories documented here are likely to vary across diverse cultural contexts (Busch, Watson-Jones, & Legare, 2018; Coley et al., 2004; Medin et al., 2010; Nielsen, Haun, Kärtner, & C. H., 2017; Ross et al., 2003). For example, most of the children in our research were growing up in urban and suburban environments, and thus perhaps receive most of their input about animal categories from picture books and other media. Such input could be comprised mainly of highly stylized exemplars, along with teleological explanations and other input that could lead them to view the categories in normative terms (or at least reinforce their normative expectations; Lee, Byatt, & Rhodes, 2000; but see also Ganea, Ma, & DeLoache, 2011). In contrast, children who grow up more closely experiencing variability within nature could incorporate more of that variability into their representations from a young age. This experience-dependent account would be consistent with other work in this area, which has found differences in category structure across children growing up with variable levels of experience with the natural world (Busch et al., 2018; Coley et al., 2004; Nielsen et al., 2017; Ross et al., 2003). For example, 5-year-old children growing up in urban environments appear to view people as the prototypical animal (Carey, 1985), whereas this is not the case for children growing up in rural environments (Medin et al., 2010).

On the other hand, it is also possible that the age-related changes documented here reflect more general developmental differences in how children represent statistical distributions or a general and pervasive focus on normative information in early childhood (phenomena that have been found across diverse cultural contexts, Diesendruck & haLevi, 2006; Nielsen & Tomaselli, 2010; Roberts, Guo, Ho, & Gelman, 2018; Taylor, 1996; Waxman, Medin, & Ross, 2007). If this is the case, then we might find more stability in these age-related trends, both across cultures as well as across domains. We suspect that ideals are especially salient for young children due to a combination of factors, including a general bias to focus on prescriptive norms (Kalish & Shiverick, 2004; Phillips & Cushman, 2017; Roberts et al., 2017). However, because the present studies focused only on animal categories, it is unclear whether the patterns observed in the current studies might be unique to biological reasoning or rather reflect more domain general features of conceptual structure. For instance, it is possible that domain-specific beliefs about category homogeneity, variability, and underlying causal mechanisms might lead to domain differences in the centrality of ideals, or in the specific sets of features viewed in idealized terms. It will be important in future work to explore the extent to which ideals might shape representations of other types of categories, including social categories like race, gender, and political affiliation. Thus, examining the extent to which the present patterns vary across diverse contexts and domains will be important for future work, both to determine the generalizability of the phenomena documented here as well to provide insight into the mechanisms underlying these age-related changes.

The current findings also do not distinguish between a preference for functionally ideal *exemplars* versus a preference for functionally ideal *properties*. Young children might value *exemplars* that they view as maximally functionally adaptive (e.g., the fastest cheetah is best able to catch its prey). However, it is at least as plausible that their choices reflect a focus on ideal *properties* as a general feature of conceptual structure (Plato, 380 B.C./1974; see also Mohr, 1977; Ziff, 1972), and that this focus manifests in the biological domain as properties that are functionally adaptive. These two interpretations make diverging predictions, for instance in the context of exemplars displaying characteristic adaptive properties to such an extreme that they become *maladaptive*—would a

giraffe whose neck is so long that it cannot be supported by its body be seen as more representative than one that is less extreme, yet more functional? Future work should carefully explore this theoretical distinction.

The findings of Study 3 suggest that people sometimes view extreme values in biological kinds as representative because they view variation in terms of progression towards a functional ideal. This does not preclude the possibility that extreme values might also sometimes take on a prominent role in biological reasoning because they are helpful for differentiation (Ameel & Storms, 2006; Davis & Love, 2010; Kim & Murphy, 2011; Levering & Kurtz, 2006; Rosch & Mervis, 1975). Although we did not find support for this possibility in Study 3, these mechanisms are not mutually exclusive. We suspect that extreme values do indeed play a greater role in representations due to contrastive value in some circumstances. Because Study 3 was designed to examine both of these mechanisms within a single paradigm (with the attention spans and memory capacities of young children in mind), we may not have set up a sufficiently sensitive test of the role of contrasting categories. For example, the contrast categories were not labeled (Waxman & Markow, 1995), they were visually similar to the target categories, and we did not measure whether participants truly represented the contrast categories as separate kinds. Thus, we interpret the pattern of findings across the various conditions of Study 3 as indicating that people do not *only* select extreme exemplars because of the contrastive mechanisms described by previous work (Ameel & Storms, 2006; Davis & Love, 2010; Levering & Kurtz, 2006)—sometimes they do so because such exemplars best instantiate functional ideals. Nevertheless, how these mechanisms work together, and especially how contrast-based mechanisms operate across development and at various stages of category learning, remain important avenues for future work.

Representative examples shape many category-based processes—they are learned more easily, better remembered, and provide more generalizable information. Documenting developmental changes in the processes underlying beliefs about representativeness is therefore crucial to our understanding of conceptual development. The current studies identified a pattern of striking developmental change in judgments of which animal category members are most representative, as well as a persistent bias in these representations towards viewing variation in terms of progression towards functional ideals. This view of the biological world can interfere with accurate scientific reasoning across the lifespan, for instance by supporting beliefs that are completely at odds with those necessary to understand evolution (e.g., by implying that variation indicates progression towards an ideal; Coley & Tanner, 2015; Kelemen, 2012; Gelman & Rhodes, 2012; Mayr, 1982; Shtulman, 2017). Thus, the current findings have significant implications, both for theories of how children learn and reason about the world and for our understanding of the nature of human thought more broadly.

## Acknowledgments

We thank John Daryl Ocampo, Amanda Cardarelli, Christine Tai, Rihab Mahmood, Christina Chu, Lauren Lucitt, Meena Jain, and Gabriela Memba for assistance with data collection, the staff and families at the American Museum of Natural History for participating in this research, Martin Wickstrom for study illustrations, and Dr. Gregory Murphy for feedback on a previous version of this manuscript.

## Appendix A. Supplementary material

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.cogpsych.2018.12.004>.

## References

- Ameel, E., & Storms, G. (2006). From prototypes to caricatures: Geometrical models for concept typicality. *Journal of Memory and Language*, 55(3), 402–421. <https://doi.org/10.1016/j.jml.2006.05.005>.
- Anglin, J. M. (1986). *Semantic and conceptual knowledge underlying the child's words*. New York, NY: Springer83–97. [http://doi.org/10.1007/978-1-4612-4844-6\\_4](http://doi.org/10.1007/978-1-4612-4844-6_4).
- Bailenson, J. N., Shum, M. S., Atran, S., Medin, D. L., & Coley, J. D. (2002). A bird's eye view: Biological categorization and reasoning within and across cultures. *Cognition*, 84(1), 1–53. [https://doi.org/10.1016/S0010-0277\(02\)00011-2](https://doi.org/10.1016/S0010-0277(02)00011-2).
- Barsalou, L. W. (1985). Ideals, central tendency, and frequency of instantiation as determinants of graded structure in categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 11(4), 629–654. <https://doi.org/10.1037/0278-7393.11.1-4.629>.
- Bjorklund, D. F., & Thompson, B. E. (1983). Category typicality effects in children's memory performance: Qualitative and quantitative differences in the processing of category information. *Journal of Experimental Child Psychology*, 35(2), 329–344. [https://doi.org/10.1016/0022-0965\(83\)90086-3](https://doi.org/10.1016/0022-0965(83)90086-3).
- Borkenau, P. (1990). Traits as ideal-based and goal-derived social categories. *Journal of Personality and Social Psychology*, 58(3), 381–396. <https://doi.org/10.1037/0022-3514.58.3.381>.
- Burnett, R. C., Medin, D. L., Ross, N. O., & Blok, S. V. (2005). Ideal is typical. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, 59(1), 3–10. <https://doi.org/10.1037/h0087453>.
- Busch, J. T. A., Watson-Jones, R. E., & Legare, C. H. (2018). Cross-cultural variation in the development of folk ecological reasoning. *Evolution and Human Behavior*, 39(3), 310–319. <https://doi.org/10.1016/j.evolhumbehav.2018.02.004>.
- Carey, S. (1985). *Conceptual change in childhood*. Cambridge, MA: Bradford Books.
- Casler, K., & Kelemen, D. (2008). Developmental continuity in teleo-functional explanation: Reasoning about nature among Romanian Romani adults. *Journal of Cognition and Development*, 9(3), 340–362. <https://doi.org/10.1080/15248370802248556>.
- Chaplin, W. F., John, O. P., & Goldberg, L. R. (1988). Conceptions of states and traits: Dimensional attributes with ideals as prototypes. *Journal of Personality and Social Psychology*, 54(4), 541–557. <https://doi.org/10.1037/0022-3514.54.4.541>.
- Cimpian, A., Brandone, A. C., & Gelman, S. A. (2010). Generic statements require little evidence for acceptance but have powerful implications. *Cognitive Science*, 34(8), 1452–1482. <https://doi.org/10.1111/j.1551-6709.2010.01126.x>.
- Coley, J. D., & Tanner, K. (2015). Relations between intuitive biological thinking and biological misconceptions in biology majors and nonmajors. *CBE Life Sciences Education*, 14(1), ar8. <https://doi.org/10.1187/cbe.14-06-0094>.
- Coley, J. D., Hayes, B., Lawson, C., & Moloney, M. (2004). Knowledge, expectations, and inductive reasoning within conceptual hierarchies. *Cognition*, 90(3), 217–253. [https://doi.org/10.1016/S0010-0277\(03\)00159-8](https://doi.org/10.1016/S0010-0277(03)00159-8).
- Costa, M., & Corazza, L. (2006). Aesthetic phenomena as supernormal stimuli: The case of eye, lip, and lower-face size and roundness in artistic portraits. *Perception*, 35(2), 229–246. <https://doi.org/10.1068/p3449>.

- Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends in Cognitive Sciences*, 13(4), 148–153. <https://doi.org/10.1016/J.TICS.2009.01.005>.
- Davis, T., & Love, B. C. (2010). Memory for category information is idealized through contrast with competing options. *Psychological Science*, 21(2), 234–242. <https://doi.org/10.1177/0956797609357712>.
- Dewar, K. M., & Xu, F. (2010). Induction, overhypothesis, and the origin of abstract knowledge. *Psychological Science*, 21(12), 1871–1877. <https://doi.org/10.1177/0956797610388810>.
- Diesendruck, G., & Gelman, S. A. (1999). Domain differences in absolute judgments of category membership: Evidence for an essentialist account of categorization. *Psychonomic Bulletin & Review*, 6(2), 338–346. <https://doi.org/10.3758/BF03212339>.
- Diesendruck, G., & haLevi, H. (2006). The role of language, appearance, and culture in children's social category-based induction. *Child Development*, 77(3), 539–553. <https://doi.org/10.1111/j.1467-8624.2006.00889.x>.
- Emmons, N. A., & Kelemen, D. A. (2015). Young children's acceptance of within-species variation: Implications for essentialism and teaching evolution. *Journal of Experimental Child Psychology*, 139, 148–160. <https://doi.org/10.1016/j.jecp.2015.05.011>.
- Ganea, P. A., Ma, L., & DeLoache, J. S. (2011). Young children's learning and transfer of biological information from picture books to real animals. *Child Development*, 82(5), 1421–1433. <https://doi.org/10.1111/j.1467-8624.2011.01612.x>.
- Gelman, S. A., & Rhodes, M. (2012). Two-thousand years of stasis. In K. S. Rosengren, S. K. Brem, E. M. Evans, & G. M. Sinatra (Eds.). *Evolution challenges: Integrating research and practice in teaching and learning about evolution* (pp. 3–21). Oxford: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199730421.003.0001>.
- Gelman, S. A., Leslie, S. J., Was, A. M., & Koch, C. M. (2015). Children's interpretations of general quantifiers, specific quantifiers and generics. *Language, Cognition and Neuroscience*, 30(4), 448–461. <https://doi.org/10.1080/23273798.2014.931591>.
- Goldstone, R. L., Steyvers, M., & Rogosky, B. J. (2003). Conceptual interrelatedness and caricatures. *Memory & Cognition*, 31(2), 169–180. <https://doi.org/10.3758/BF03194377>.
- Gutheil, G., & Gelman, S. A. (1997). Children's use of sample size and diversity information within basic-level categories. *Journal of Experimental Child Psychology*, 64(2), 159–174.
- Haward, P., Wagner, L., Carey, S., & Prasada, S. (2018). The development of principled connections and kind representations. *Cognition*, 176, 255–268. <https://doi.org/10.1016/j.cognition.2018.02.001>.
- Heit, E., & Feeney, A. (2005). Relations between premise similarity and inductive strength. *Psychonomic Bulletin & Review*, 12(2), 340–344.
- Heit, E., & Hahn, U. (2001). Diversity-based reasoning in children. *Cognitive Psychology*, 43(4), 243–273. <https://doi.org/10.1006/COGP.2001.0757>.
- Heit, E., Hahn, U., & Feeney, A. (2004). Defending diversity. In W. Ahn, R. L. Goldstone, B. C. Love, A. B. Markman, & P. Wolff (Eds.). *Categorization inside and outside the lab: Festschrift in honor of Douglas L. Medin* (pp. 87–99). Washington, DC: American Psychological Association.
- Horst, J. S., & Hout, M. C. (2016). The Novel Object and Unusual Name (NOUN) Database: A collection of novel images for use in experimental research. *Behavior Research Methods*, 48(4), 1393–1409. <https://doi.org/10.3758/s13428-015-0647-3>.
- Judd, C. M., Westfall, J., & Kenny, D. A. (2012). Treating stimuli as a random factor in social psychology: A new and comprehensive solution to a pervasive but largely ignored problem. *Journal of Personality and Social Psychology*, 103(1), 54–69. <https://doi.org/10.1037/a0028347>.
- Kalish, C. (1998). Reasons and causes: Children's understanding of conformity to social rules and physical laws. *Child Development*, 69(3), 706–720. <https://doi.org/10.1111/j.1467-8624.1998.tb06238.x>.
- Kalish, C. W., & Lawson, C. A. (2008). Development of social category representations: Early appreciation of roles and deontic relations. *Child Development*, 79(3), 577–593. <https://doi.org/10.1111/j.1467-8624.2008.01144.x>.
- Kalish, C. W., & Shiverick, S. M. (2004). Children's reasoning about norms and traits as motives for behavior. *Cognitive Development*, 19(3), 401–416. <https://doi.org/10.1016/J.COGEDEV.2004.05.004>.
- Kelemen, D. (1999). Why are rocks pointy? Children's preference for teleological explanations of the natural world. *Developmental Psychology*, 35(6), 1440–1452. <https://doi.org/10.1037/0012-1649.35.6.1440>.
- Kelemen, D. (2003). British and American children's preferences for teleo-functional explanations of the natural world. *Cognition*, 88(2), 201–221. [https://doi.org/10.1016/S0010-0277\(03\)00024-6](https://doi.org/10.1016/S0010-0277(03)00024-6).
- Kelemen, D. (2012). Teleological minds: How natural intuitions about agency and purpose influence learning about evolution. In K. S. Rosengren, S. K. Brem, E. M. Evans, & G. M. Sinatra (Eds.). *Evolution challenges: Integrating research and practice in teaching and learning about evolution* (pp. 66–92). Oxford: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199730421.003.0004>.
- Kelemen, D., Callanan, M. A., Casler, K., & Pérez-Granados, D. R. (2005). Why things happen: Teleological explanation in parent-child conversations. *Developmental Psychology*, 41(1), 251–264. <https://doi.org/10.1037/0012-1649.41.1.251>.
- Kelemen, D., Rottman, J., & Seston, R. (2013). Professional physical scientists display tenacious teleological tendencies: Purpose-based reasoning as a cognitive default. *Journal of Experimental Psychology: General*, 142(4), 1074–1083. <https://doi.org/10.1037/a0030399>.
- Kim, N. S., & Keil, F. C. (2003). From symptoms to causes: Diversity effects in diagnostic reasoning. *Memory & cognition*, 31(1), 155–165. <https://doi.org/10.3758/BF03196090>.
- Kim, S., & Murphy, G. L. (2011). Ideals and category typicality. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37(5), 1092–1112. <https://doi.org/10.1037/a0023916>.
- Lee, K., Byatt, G., & Rhodes, G. (2000). Caricature effects, distinctiveness, and identification: Testing the face-space framework. *Psychological Science*, 11(5), 379–385. <https://doi.org/10.1111/1467-9280.00274>.
- Levering, K., & Kurtz, K. J. (2006). The influence of learning to distinguish categories on graded structure. In R. Sun, & N. Miyake (Eds.). *Proceedings of the 28th annual conference of the Cognitive Science Society* (pp. 1681–1686). Austin, TX: Cognitive Science Society.
- Lombrozo, T., Kelemen, D., & Zaitchik, D. (2007). Inferring design. *Psychological Science*, 18(11), 999–1006. <https://doi.org/10.1111/j.1467-9280.2007.02015.x>.
- López, A., Gelman, S. A., Gutheil, G., & Smith, E. E. (1992). The development of category-based induction. *Child Development*, 63, 1070–1090. <https://doi.org/10.2307/1131519>.
- Lynch, E. B., Coley, J. D., & Medin, D. L. (2000). Tall is typical: Central tendency, ideal dimensions, and graded category structure among tree experts and novices. *Memory & Cognition*, 28(1), 41–50. <https://doi.org/10.3758/BF03211575>.
- Mayr, E. (1982). *The growth of biological thought: Diversity, evolution, and inheritance*. Harvard University Press.
- Medin, D. L., Wattenmaker, W. D., & Hampson, S. E. (1987). Family resemblance, conceptual cohesiveness, and category construction. *Cognitive Psychology*, 19(2), 242–279. [https://doi.org/10.1016/0010-0285\(87\)90012-0](https://doi.org/10.1016/0010-0285(87)90012-0).
- Medin, D., Waxman, S., Woodring, J., & Washinawatok, K. (2010). Human-centeredness is not a universal feature of young children's reasoning: Culture and experience matter when reasoning about biological entities. *Cognitive Development*, 25(3), 197–207. <https://doi.org/10.1016/J.COGEDEV.2010.02.001>.
- Mervis, C. B., & Pani, J. R. (1980). Acquisition of basic object categories. *Cognitive Psychology*, 12(4), 496–522. [https://doi.org/10.1016/0010-0285\(80\)90018-3](https://doi.org/10.1016/0010-0285(80)90018-3).
- Mohr, R. D. (1977). Family resemblance, Platonism, universals. *Canadian Journal of Philosophy*, 7(3), 593–600. <https://doi.org/10.2307/40230707>.
- Murphy, G. L. (2002). *The big book of concepts*. MIT Press 555.
- Murphy, G. L., & Brownell, H. H. (1985). Category differentiation in object recognition: Typicality constraints on the basic category advantage. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 11, 70–84. <https://doi.org/10.1037/0278-7393.11.1.70>.
- Nielsen, M., Haun, D., Kärtner, J., & Legare, C. H. (2017). The persistent sampling bias in developmental psychology: A call to action. *Journal of Experimental Child Psychology*, 162, 31–38. <https://doi.org/10.1016/J.JECP.2017.04.017>.
- Nielsen, M., & Tomaselli, K. (2010). Overimitation in Kalahari Bushman children and the origins of human cultural cognition. *Psychological Science*, 21(5), 729–736. <https://doi.org/10.1177/0956797610368808>.
- Osherson, D. N., Smith, E. E., Wilkie, O., López, A., & Shafir, E. (1990). Category-based induction. *Psychological Review*, 97(2), 185–200. <https://doi.org/10.1037/0033-295X.97.2.185>.
- Phillips, J., & Cushman, F. (2017). Morality constrains the default representation of what is possible. *Proceedings of the National Academy of Sciences of the United States*



- of America, 114(18), 4649–4654. <https://doi.org/10.1073/pnas.1619717114>.
- Plato, A., & Grube, G. M. A. (1974). *The Republic*. Hackett Publishing.
- Prasada, S., & Dillingham, E. M. (2006). Principled and statistical connections in common sense conception. *Cognition*, 99(1), 73–112. <https://doi.org/10.1016/J.COgnITION.2005.01.003>.
- Prasada, S., & Dillingham, E. M. (2009). Representation of principled connections: A window onto the formal aspect of common sense conception. *Cognitive Science*, 33(3), 401–448. <https://doi.org/10.1111/j.1551-6709.2009.01018.x>.
- Rakoczy, H., & Schmidt, M. F. H. (2013). The early ontogeny of social norms. *Child Development Perspectives*, 7(1), 17–21. <https://doi.org/10.1111/cdep.12010>.
- Read, S. J., Jones, D. K., & Miller, L. C. (1990). Traits as goal-based categories: The importance of goals in the coherence of dispositional categories. *Journal of Personality and Social Psychology*, 58(6), 1048–1061. <https://doi.org/10.1037/0022-3514.58.6.1048>.
- Rein, J. R., Goldwater, M. B., & Markman, A. B. (2010). What is typical about the typicality effect in category-based induction? *Memory & Cognition*, 38(3), 377–388. <https://doi.org/10.3758/MC.38.3.377>.
- Rhodes, M., Bonawitz, E., Shafto, P., Chen, A., & Caglar, L. (2015). Controlling the message: Preschoolers' use of information to teach and deceive others. *Frontiers in Psychology*, 6, 867. <https://doi.org/10.3389/fpsyg.2015.00867>.
- Rhodes, M., Brickman, D., & Gelman, S. A. (2008). Sample diversity and premise typicality in inductive reasoning: Evidence for developmental change. *Cognition*, 108(2), 543–556. <https://doi.org/10.1016/J.COgnITION.2008.03.002>.
- Rhodes, M., & Liebensohn, P. (2015). Continuity and change in the development of category-based induction: The test case of diversity-based reasoning. *Cognitive Psychology*, 82, 74–95. <https://doi.org/10.1016/J.COgnPSYCH.2015.07.003>.
- Rips, L. J. (1975). Inductive judgments about natural categories. *Journal of Verbal Learning and Verbal Behavior*, 14(6), 665–681. [https://doi.org/10.1016/S0022-5371\(75\)80055-7](https://doi.org/10.1016/S0022-5371(75)80055-7).
- Rips, L. J., Shoben, E. J., & Smith, E. E. (1973). Semantic distance and the verification of semantic distance. *Journal of Verbal Learning and Verbal Behavior*, 12, 1–20.
- Roberts, S. O., Guo, C., Ho, A. K., & Gelman, S. A. (2018). Children's descriptive-to-prescriptive tendency replicates (and varies) cross-culturally: Evidence from China. *Journal of Experimental Child Psychology*, 165, 148–160. <https://doi.org/10.1016/J.JECP.2017.03.018>.
- Roberts, S. O., Ho, A. K., & Gelman, S. A. (2017). Group presence, category labels, and generic statements influence children to treat descriptive group regularities as prescriptive. *Journal of Experimental Child Psychology*, 158, 19–31. <https://doi.org/10.1016/J.JECP.2016.11.013>.
- Rosch, E. H. (1973). Natural categories. *Cognitive Psychology*, 4(3), 328–350. [https://doi.org/10.1016/0010-0285\(73\)90017-0](https://doi.org/10.1016/0010-0285(73)90017-0).
- Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, 7(4), 573–605. [https://doi.org/10.1016/0010-0285\(75\)90024-9](https://doi.org/10.1016/0010-0285(75)90024-9).
- Rosch, E., Simpson, C., & Miller, R. S. (1976). Structural bases of typicality effects. *Journal of Experimental Psychology: Human Perception and Performance*, 2(4), 491–502. <https://doi.org/10.1037/0096-1523.2.4.491>.
- Rosner, S. R., & Hayes, D. S. (1977). A developmental study of category item production. *Child Development*, 48(3), 1062. <https://doi.org/10.2307/1128361>.
- Ross, N., Medin, D., Coley, J. D., & Atran, S. (2003). Cultural and experiential differences in the development of folkbiological induction. *Cognitive Development*, 18(1), 25–47. [https://doi.org/10.1016/S0885-2014\(02\)00142-9](https://doi.org/10.1016/S0885-2014(02)00142-9).
- Shafto, P., Goodman, N. D., & Griffiths, T. L. (2014). A rational account of pedagogical reasoning: Teaching by, and learning from, examples. *Cognitive Psychology*, 71, 55–89. <https://doi.org/10.1016/J.COgnPSYCH.2013.12.004>.
- Shtulman, A. (2006). Qualitative differences between naïve and scientific theories of evolution. *Cognitive Psychology*, 52(2), 170–194. <https://doi.org/10.1016/j.cogpsych.2005.10.001>.
- Shtulman, A. (2017). *Scienceblind: Why our intuitive theories about the world are so often wrong*. Basic Books.
- Shtulman, A., & Carey, S. (2007). Improbable or impossible? How children reason about the possibility of extraordinary events. *Child Development*, 78(3), 1015–1032. <https://doi.org/10.1111/j.1467-8624.2007.01047.x>.
- Shtulman, A., & Harrington, K. (2016). Tensions between science and intuition across the lifespan. *Topics in Cognitive Science*, 8(1), 118–137. <https://doi.org/10.1111/tops.12174>.
- Shtulman, A., & Phillips, J. (2018). Differentiating “could” from “should”: Developmental changes in modal cognition. *Journal of Experimental Child Psychology*, 165, 161–182. <https://doi.org/10.1016/j.jecp.2017.05.012>.
- Shtulman, A., & Schulz, L. (2008). The relation between essentialist beliefs and evolutionary reasoning. *Cognitive Science*, 32(6), 1049–1062. <https://doi.org/10.1080/03640210801897864>.
- Smith, J. D., & Minda, J. P. (1998). Prototypes in the mist: The early epochs of category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24(6), 1411–1436. <https://doi.org/10.1037/0278-7393.24.6.1411>.
- Springer, K., & Keil, F. C. (1989). On the development of biologically specific beliefs: The case of inheritance. *Child Development*, 637–648. <https://doi.org/10.2307/1130729>.
- Taylor, M. G. (1996). The development of children's beliefs about social and biological aspects of gender differences. *Child Development*, 67(4), 1555. <https://doi.org/10.2307/1131718>.
- Tisak, M. S., & Turiel, E. (1988). Variation in seriousness of transgressions and children's moral and conventional concepts. *Developmental Psychology*, 24(3), 352–357. <https://doi.org/10.1037/0012-1649.24.3.352>.
- Ware, E. A., & Gelman, S. A. (2014). You get what you need: An examination of purpose-based inheritance reasoning in undergraduates, preschoolers, and biological experts. *Cognitive Science*, 38(2), 197–243. <https://doi.org/10.1111/cogs.12097>.
- Waxman, S. R., & Markow, D. B. (1995). Words as invitations to form categories: Evidence from 12- to 13-month-old infants. *Cognitive Psychology*, 29(3), 257–302. <https://doi.org/10.1006/COGP.1995.1016>.
- Waxman, S., Medin, D., & Ross, N. (2007). Folkbiological reasoning from a cross-cultural developmental perspective: Early essentialist notions are shaped by cultural beliefs. *Developmental Psychology*, 43(2), 294.
- Xu, F., & Tenenbaum, J. B. (2007). Word learning as Bayesian inference. *Psychological Review*, 114(2), 245–272. <https://doi.org/10.1037/0033-295X.114.2.245>.
- Zhong, L., Lee, M. S., Huang, Y., & Mo, L. (2014). Diversity effect in category-based inductive reasoning of young children: Evidence from two methods. *Psychological Reports*, 114(1), 198–215.
- Ziff, P. (1972). *Understanding understanding*. Ithaca: Cornell University Press.