Developmental Psychology

The Development of the “First Thing That Comes to Mind”

Jonathan F. Kominsky, Joshua Knobe, and Elizabeth Bonawitz

Online First Publication, June 5, 2025. https://dx.doi.org/10.1037/dev0001994

CITATION

Kominsky, J. F., Knobe, J., & Bonawitz, E. (2025). The development of the “first thing that comes to mind”. *Developmental Psychology*. Advance online publication. https://dx.doi.org/10.1037/dev0001994

AMERICAN PSYCHOLOGICAL ASSOCIATION

Developmental Psychology

ISSN: 0012-1649

https://doi.org/10.1037/dev0001994

The Development of the "First Thing That Comes to Mind"

Jonathan F. Kominsky[1], Joshua Knobe[2], and Elizabeth Bonawitz[3]
[1] Department of Cognitive Science, Central European University
[2] Department of Philosophy, Program in Cognitive Science, Yale University
[3] Graduate School of Education, Harvard University

When children encounter information about the world that is descriptive (e.g., frequency distributions) or prescriptive (e.g., value judgments), can they keep track of both types of information together? Do they, like adults, integrate these two kinds of information to come up with the "first thing that comes to mind"? Can children separate these types of information when needed? In two experiments, we examined how children ($N = 397$, ages 4–9 years, fluent English speakers mostly from North America, recruited online) and adults ($N = 189$, U.S. English speakers, recruited online) produce both "first-to-mind" judgments and predictions about random samples. In Experiment 1, providing information about whether being longer or shorter made a fictional tool better or worse led adults to provide first-to-mind judgments that were biased toward the prescriptive ideal, but unbiased random sample predictions. However, 6–9-year-old children provided judgments that were biased by the prescriptive ideal in both cases. In Experiment 2, with 6–9-year-olds and adults, we manipulated whether the prescriptive information focused exclusively on positive (i.e., only "better") or negative (i.e., only "worse") properties. In the positive-focus condition, all age groups showed an effect of prescriptive ideal on first-to-mind judgments, but only 6–7-year-olds showed an effect of prescriptive ideal on random sample predictions. However, in the negative-focus condition, there was no effect of prescriptive information on either type of judgments for any age group, including adults. We discuss what changes in development in the ability to represent different kinds of information and apply the best kind of information to a specific task.

Public Significance Statement

We show that children, like adults, keep track of both how likely something is and whether it is "good." However, 6–7-year-olds also use information about what is "good" in cases where adults only use information about what is likely, for example when making predictions about random events. This tells us children in this age range might need help knowing what kind of information to use for different tasks, even though they can keep track of different kinds of information.

Keywords: cognitive development, decision making, sampling processes

Supplemental materials: https://doi.org/10.1037/dev0001994.supp

Emily Ruth Fyfe served as action editor.

Jonathan F. Kominsky https://orcid.org/0000-0002-3236-4787

Experiment 1 materials, data, and analysis scripts for the adult pilot can be found on the Open Science Framework at https://osf.io/9x6y2/ and for the developmental population at https://osf.io/9xu8t/. Experiment 2 materials, data, and analysis scripts can be found at https://osf.io/nrjaz/. OSF repository at https://osf.io/64uqk/. The authors thank Elise Mahaffey and Michelle Wong for their help with recruitment. Jonathan F. Kominsky was supported by a John Templeton Foundation Developing Belief Network postdoctoral fellowship. Elizabeth Bonawitz was supported by James S. McDonnell Foundation, Inc. (Grant 220020544). This publication is the result of research conducted for Central European University, Private University. It was made possible by the Central European University Open Access Fund.

Jonathan F. Kominsky contributed to conceptualization, data curation, formal analysis, investigation, methodology, software, visualization, writing–original draft, and writing–review and editing. Joshua Knobe contributed to conceptualization, supervision, and writing–review and editing. Elizabeth Bonawitz contributed to conceptualization, supervision, resources, and writing–review and editing.

Jonathan F. Kominsky played a lead role in data curation, formal analysis, investigation, methodology, software, visualization, writing–original draft, and writing–review and editing and an equal role in conceptualization. Joshua Knobe played a supporting role in supervision and an equal role in conceptualization and writing–review and editing. Elizabeth Bonawitz played a lead role in resources and supervision and an equal role in conceptualization and writing–review and editing.

Correspondence concerning this article should be addressed to Jonathan F. Kominsky, Department of Cognitive Science, Central European University, 51–55 Quellenstrasse, 1100 Vienna, Austria. Email: kominskyj@ceu.edu

As developmental psychologists, one of the central questions we focus on is how people acquire and make use of representations of the world over development. Adults have the ability to think about different kinds of information in relation to a single event or type of object, such as descriptive information extracted from our observation of that type of thing in the world, and prescriptive information determined not by what we have observed, but by what we would like to be the case, or what our culture or society has determined should be the case. Moreover, adults can make use of these different kinds of information for different tasks and, in some cases, synthesize them. For example, consider a simple question: How long is a conference poster session? We can describe whatever duration just popped into your head as the "first thing that came to mind" for this feature of your experience. Now ask yourself two further questions: What do you think is the ideal length for a conference poster session? What is the average length of the conference poster sessions that you have attended? Intuitively, one might think that the "first thing that comes to mind" would simply be one of these two (i.e., entirely based on descriptive information or entirely based on prescriptive information), but you may find that all three have different values.

Past work has found that the "first thing to mind" for adults in cases like the example above tends to be a combination of the descriptive average of a probability distribution of what the respondent has observed and the prescriptive ideal of what they think would be best (Bear et al., 2020). This seems to be deeply related to two other common concepts in psychology. The first related idea is judgments of what is "normal"; e.g., the "normal" number of hours of television to watch in a day is also a combination of the average and the ideal (Bear & Knobe, 2017; Phillips et al., 2019). The second related concept is what is "typical" of a category; for example, a tool that is the best at fulfilling its goal is often judged to be more typical than a tool with average performance (Barsalou, 1985; cf. Kim & Murphy, 2011). Notably, adults can report the average and the ideal separately, suggesting that they have an internal representation that tracks both a distribution of frequency and a distribution of prescriptive value, or the ability to decompose the representation that drives their "first-to-mind" intuitions into its component parts.

All of the work on "first-to-mind" judgments to date has focused on adults and suggested a complex and sophisticated set of representations and processes for these nominally "intuitive" judgments. However, these accounts raise some notable questions about how these judgments might work differently in children, particularly since the relationship between descriptive and prescriptive information in children seems to work differently than it does in adults. Understanding this relationship and how it arises in development speaks to our broader theories of what factors influence how information about the world is represented in early childhood and how those representations might change with age.

Prescriptive and Descriptive Knowledge in Development

In the absence of prescriptive information, children demonstrate sensitivity to the statistics of the environment. It is well-established that even infants are very adept at tracking probabilistic information. By 8 months of age, infants predict the result of a random sampling event by the observed distribution of outcomes (Xu & Denison, 2009; Xu & Garcia, 2008), and by 11 months, they are able to anticipate the outcome of probabilistic events (Téglás & Bonatti, 2016). By 4–5 years of age, children have even more sophisticated intuitions about probabilistic events, providing responses that reflect sampling from posterior distributions (Denison et al., 2013). That is, although any individual child's response many look like a noisy or random sample, children's responses as a group reflect the distribution they are representing, such that if (e.g.,) a distribution was split 70/30 for Options A and B, respectively, about 70% of children would answer "A" and 30% would answer "B."

Furthermore, preschoolers integrate some inductive biases (such as a preference for simpler explanations) with statistical information (such as the conditional probability of events) in their spontaneous causal inferences (Bonawitz & Lombrozo, 2012). That is, if two explanations are equally statistically likely to occur, but one explanation is more complex (requires more causal elements), children are more likely to prefer the simpler explanation. As the complex explanation becomes more statistically likely than the simple one, children will start to trade off this preference for simplicity with the evidence and select the more complex answer more often.

Notably, an underlying representation of a frequency distribution allows adults to calculate an average, but the concept of an average or mean is difficult to explain to a child. However, making accurate predictions about random sampling events requires sampling from that same kind of frequency distribution. If children represent the underlying frequency distribution accurately, then these predictions should, in aggregate, recreate the original distribution. In short, asking children to predict a token random sampling event from a distribution demonstrates the same capacity (albeit potentially implicitly) to represent descriptive frequency that we find in adults when asking them to calculate an average.

Although they are savvy statistical reasoners, the literature has suggested that children frequently confound descriptive and prescriptive information. For example, up to 10 years of age, children will call atypical or unlikely behavior morally wrong (Tisak & Turiel, 1988), and up to age 6, they will often say that immoral events are actually physically impossible (Shtulman & Phillips, 2018), just as they do for many unlikely (but physically possible) events (Shtulman & Carey, 2007). In other words, children may not make a distinction between descriptive and prescriptive information in the first place. Notably, some of the consequences of this developmental confusion seem to persist into adulthood, in that adults sometimes seem to develop intuitions about how things *ought* to be from the way things currently *are* (Tworek & Cimpian, 2016).

Further evidence of children's (and adults') tendency to mix prescriptive and descriptive information comes from the extensive literature on category representations and prototypes. On one hand, children and adults use the descriptive frequency of a distinctive trait to evaluate whether a member of a category "ought" to have that trait (Foster-Hanson & Lombrozo, 2022). Conversely, when human and nonhuman animals engage in category-atypical behavior, children judge it to be prescriptively wrong (Foster-Hanson et al., 2021; Roberts et al., 2017).

Interestingly, when asked to pick an example of a category of animal to show to someone who has never seen one before, 5–6-year-old children tend to pick one that shows a rare but extreme value of that animal's distinctive trait (e.g., the fastest cheetah), while older children and adults tend to pick one that represents integrating both average and informative features (e.g., one that has more than the average value of the trait, but less than the maximum).

At the same time, when asked directly for the "best" member of a category, both children and adults have no difficulty selecting the one with the most extreme value (Foster-Hanson & Rhodes, 2019).[1] Most strikingly, when children are asked to pick out the "real cheetah-y cheetah" or the "real chair-y chair," their responses are impacted by prescriptive information (Foster-Hanson & Rhodes, 2019). Thus, children's choice of the real cheetah-y cheetah ends up being not the statistical central tendency but rather a cheetah that is especially fast. This research seems especially directly relevant to questions about which category exemplars will be first to mind for children.

What, then, is the relationship between children's ability to track descriptive regularities in the world, their intuitions about prescriptive value, their "first-to-mind" judgments, and their ability to predict token random sampling events? Given that they seem to confuse prescriptive and descriptive information anyway, one obvious prediction is that children's first-to-mind judgments would align with one or the other. However, the aforementioned work on selecting exemplars of a category might indicate that children instead employ a combination of the two. We can frame these as three distinct hypotheses about children's representations of descriptive information, prescriptive information, and the method by which they make "first-to-mind" judgments and predictions about token random sampling events (see also Figure 1):

Hypothesis 1: "Prescriptive bias": Prescriptive information is prioritized over all other information. Therefore, children's "first-to-mind" judgments will be strongly influenced by the prescriptive ideal or even identical to it. Furthermore, when asked for a prediction of a token random sample event, their response will be strongly biased in the direction of the prescriptive ideal. Over development, they incorporate descriptive information from their observations of the world and are able to use representations of a frequency distribution for their random sample predictions, as well as incorporating descriptive information into their first-to-mind judgments.

Hypothesis 2: "Frequency wins": Descriptive frequency is the primary thing that children track, and they infer prescriptive value from the assumption that what is frequent is also good. When asked to predict a token random sample event, they will do so accurately, that is, several of these tokens in aggregate will reproduce the underlying frequency distribution. In addition, children's "first-to-mind" judgments will be identical to random samples from the distribution of observed frequencies. Over development, they begin to incorporate the prescriptive ideal into their first-to-mind judgments.

Hypothesis 3: "Undifferentiated information": Children do not distinguish between prescriptive and descriptive information, but rather represent any given feature of the world as a combination of the two. Like adults, children's first-to-mind judgments will be a combination of descriptive frequency and prescriptive ideal, but unlike adults, when asked to predict a random sample, children's responses will *also* be influenced by the prescriptive ideal. Over development, children will learn to distinguish descriptive frequency and prescriptive value and be able to report each value separately, but their "first-to-mind" judgments will remain a combination of the two, drawn from this underlying undifferentiated representation.

In two experiments, we aimed to distinguish between these three hypotheses by building on a task used by Bear et al. (2020; itself based on a task in Bear & Knobe, 2017). We presented participants with an artificial scenario and a distribution with descriptive and prescriptive characteristics that we could precisely control. Experiment 1 adapts Bear et al. (2020)'s Experiment 3 to be feasible with children ages 4–9 years, while also attempting to replicate the earlier article's findings with adults. We presented participants with 100 examples of each of two novel types of objects that each varied in a single continuous feature dimension (e.g., length). The 100 examples were the product of sampling from a normal distribution of the feature dimension (e.g., average length = 40). These examples were presented alongside information about the prescriptive value of objects that varied with this feature dimension (e.g., "longer is better"). Participants were asked to either report the first object that came to mind or predict a random sample from the distribution they observed.

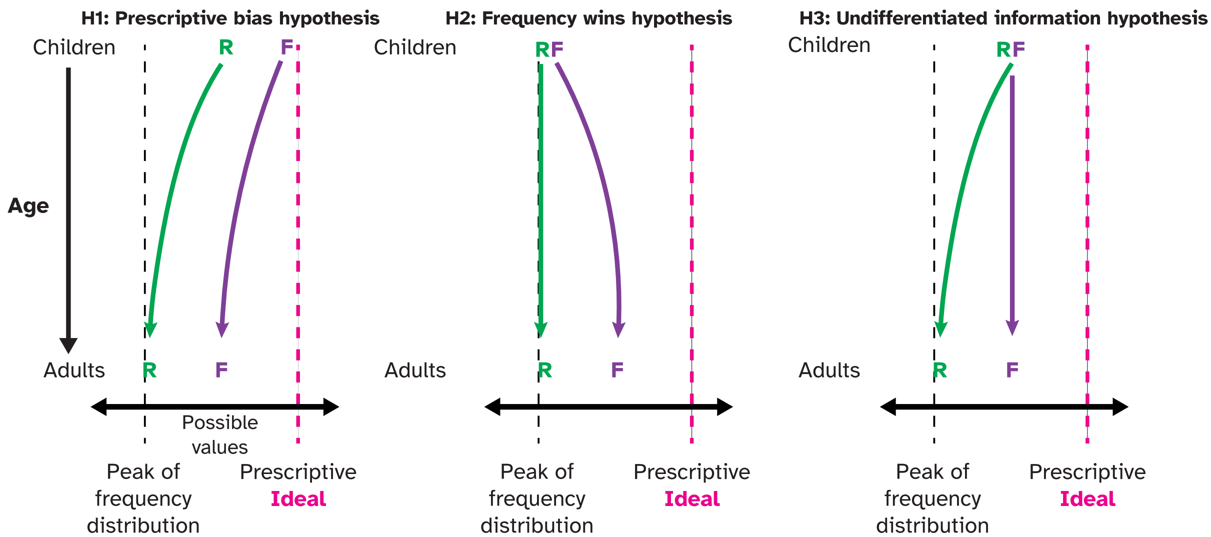
For adults, we predicted that (replicating past work) first-to-mind judgments would reflect a combination of the peak of the frequency distribution and the prescriptive ideal and that random sample judgments would in aggregate recreate the frequency distribution they were shown. The "Prescriptive bias" hypothesis predicts that children's judgments will be more strongly affected by the prescriptive ideal relative to adults' judgments, regardless of the question. The "Frequency wins" hypothesis predicts that children and adults will perform similarly on random sample predictions, but that children's first-to-mind judgments will also be samples from the observed frequency distribution (i.e., be unaffected by the prescriptive ideal). The "Undifferentiated information" hypothesis predicts that children and adults will provide similar first-to-mind judgments, but that children's random sample predictions will also look like their (and adults') first-to-mind judgments, that is, their random samples will also be influenced by prescriptive ideal, and to a similar degree as their first-to-mind judgments.

Notably, for the "Prescriptive bias" and "Undifferentiated information" hypotheses, there are different ways that prescriptive information could influence children's responses (and, for that matter, how it could influence adults'). Previous work, and Experiment 1, either presented only positive prescriptive information or both positive and negative prescriptive information together, so it is unknown whether the influence of prescriptive information is driven by focusing on things that are prescriptively good, avoiding things that are prescriptively bad, or a combination of the two. Experiment 2 was designed to both control for a deflationary explanation of children's responses in Experiment 1 and to evaluate the role of positive versus negative prescriptive information on both children's and adults' first-to-mind judgments and their predictions about random sampling events.

[1] While older children and adults' selected example seems very similar to adults' first-to-mind judgments (in that it typically falls between the average and the ideal), it is not quite the same judgment. In particular, it is a selection to fulfill a particular goal (introducing the category to someone who is unaware of it), which may reflect deliberative reasoning about pedagogical goals and what the audience might know (something that children keep track of by ages 5–6; e.g., Bass et al., 2019, 2022, 2023; Rhodes et al., 2015) rather than being simply the first example that came to mind.

Figure 1

Three Hypotheses About the Development of "First-to-Mind" Judgments, F (Purple) in Relation to "Random Sample" Judgments, R, (Green)



Note. The left dashed line indicates the value corresponding to the true peak of a normal frequency distribution, while the right dashed line (in magenta) indicates a prescriptively ideal value. The letters at the top represent children's judgments, while those at the bottom represent adults', reflecting the developmental trajectory. The more the letters are shifted to the right, the more those judgments are influenced by the prescriptive ideal. See text for explanation of each hypothesis. H = hypothesis. See the online article for the color version of this figure.

Experiment 1

In our first experiment, our goal was twofold. First, to replicate previous results with adults showing that prescriptive information influences their first-to-mind judgments, but not predictions about random samples from a distribution. Second, to look at the behavior of 4–9-year-old children and test the Prescriptive bias, Frequency wins, and Undifferentiated information hypotheses.

Method

Transparency and openness

The developmental sample in Experiment 1 was preregistered (https://osf.io/nu8vm; updated registration at https://osf.io/pcwzn), and the adult pilot was registered while data collection was ongoing but prior to analysis (https://osf.io/jpazm). All materials, data, and analysis scripts for the adult pilot can be found at https://osf.io/9x6y2/ and for the developmental population at https://osf.io/9xu8t/.

Participants

We first ran a study (not preregistered) with adults, aiming to collect 40 responses in each Question condition (first-to-mind vs. random sample prediction) that passed the exclusion criteria (see below). Our final sample consisted of 80 adults (34 female, 45 male, one declined to specify) recruited from Prolific who passed criteria, and an additional three adults (one female, two male) who were excluded and replaced prior to analysis.

Based on a power analysis of the results of this pilot, we preregistered (https://osf.io/nu8vm; https://osf.io/pcwzn) a plan to recruit a minimum of 52 children in each child age group (4–5-year-olds, 6–7-year-olds, 8–9-year-olds), split evenly between two Question conditions (26 first-to-mind vs. 26 random sample) who passed the exclusion check items (see below). However, because of the nature of unmoderated data collection, we said we would continue collecting data until we had *at least* this many participants in each group and also include the data from any participants who participated before the last group reached the target minimum if they passed the check items. We recruited by advertising on Children Helping Science (Sheskin et al., 2020) prior to its merger with Lookit (see Experiment 2).

Our final sample consisted of 59 4- and 5-year-olds (26 in the random sample condition, 33 in the first-to-mind condition, 23 female, 36 male), 63 6- and 7-year-olds (33 in the random sample condition, 30 in the first-to-mind condition, 29 female, 33 male, one did not specify), and 56 8- and 9-year-olds (29 in the random sample condition, 27 in the first-to-mind condition, 24 female, 31 male, one did not specify). An additional 42 4- and 5-year-olds (21 female, 21 male), 28 6- and 7-year-olds (11 female, 17 male), and 20 8- and 9-year-olds (five female, 14 male, one did not specify) were excluded per preregistered criteria (see below) and replaced prior to analyses.

Stimuli and Procedure

The study was presented as a website, custom-programmed in PHP and JavaScript. PHP was used for condition assignment and data recording, while JavaScript controlled the actual stimulus presentation and response interface. The code can be found at the OSF repository at https://osf.io/9xu8t/.

Participants or their parents first completed a consent form and filled in demographic data (gender identity and age in years). Because the consent data and the response data were saved together (due to the design of the webpage itself), we could not ask for any potentially identifying information (e.g., race/ethnicity or exact date of birth), so we do not have detailed demographic data for this experiment (this was remedied by using a separate webpage for consent in Experiment 2). Study sessions were not video- or audio-recorded. After filling out the consent and demographics, participants then completed a brief technical validation to ensure that their browser could handle the relevant JavaScript libraries. In particular, the technical validation tested the primary response method, which was a circular slider. The slider appeared as a circle 580 pixels in diameter with a 20 pixel by 30 pixel blue rectangle on it. As participants moved their mouse around the screen, the slider moved around the circle (changing orientation appropriately). When participants clicked, the rectangle turned yellow and locked in place. The rectangle could be clicked again to unlock it in order to change a response. This circular slider was used for all of the primary dependent variables throughout this experiment. For the technical test, inside the circle was a smaller filled-in black circle that grew and shrank as the slider moved around the outer circle. Participants had to correctly select the shape of the inner circle and say that it changed size as the slider moved around in order to continue.

Following these preliminaries, participants saw an instructions page informing them that they were going to play a game where they would be asked different kinds of questions, and first, we would introduce them to the different questions we might ask them. All instructions were presented in text form along with a prerecorded narration that matched the written text exactly. Participants could not advance to the next page until the audio narration had finished.

Participants then completed three training items in a randomized order. One training item asked them to find the "best thing." The example used was a test where one could get up to four stars if they got all the questions right. Participants were shown images of papers with 1–4 stars and asked to pick the "best" outcome, which was the article with four stars. If they picked another item, they were told this was incorrect and told to try to pick the "best" item again. Then the images were replaced with the circular slider described above. The image in the center of the circular slider consisted of a number of stars between 1 and 4, and participants were instructed to use the slider to make it show four stars. They were not allowed to proceed until they had done so. If they selected an incorrect number of stars at either step, they heard an additional audio recording telling them that was not right and they should try again.

Another training item told them they might be asked to predict something. They were shown an image of a box with 28 balls in it, 26 of which were green, one of which was orange, and one of which was blue. The image then showed the box being shaken and a hand pushing a button on the top of the box with a chute coming out of the box and a question mark. Participants were told that when someone pushes the button, a random ball comes out and were asked to predict what color of ball would come out when someone pushed the button. They could select an orange, blue, or green ball. If they selected something other than the green ball, they were prompted to choose again. When they selected the green ball, the images were replaced by the circular slider, and they were asked to make the circle inside the slider match the color they selected. The slider in this item corresponded to a circular HSL (Hue/Saturation/Lightness) color space, and participants had to make the color of the central circle within ±50° of the green used in the image before they could proceed. Thus, on these two training items, they had to demonstrate both that they understood the question and that they could use the slider to produce the correct answer before they could proceed to the main task.

The other training item told them they might be asked to use the slider to report "the first thing that comes to mind." For this item, they were asked to think of a color, then type the name of that color in a small text box or ask their parent to type it for them. Then they were told to use the slider (using the same HSL color space as the prediction training item) to make that color. This item had no validation.

After completing the training, participants moved on to the first learning phase. Participants were told about an object called a "Dax" with an image of a spear-like object. They read and saw the following text:

> We found some aliens on planet Debian! On Debian, they make tools called Daxes for catching fish. Here's an example of a Dax. Some Daxes are better at catching fish than others. Some are awful, some are very bad, some are a little bad, some are a little good, others are very good, and some are awesome.

Then, they were told:

> The awesome Daxes that can catch the most fish are the [longest/shortest] Daxes, while the awful Daxes that catch the fewest fish are the [shortest/longest] Daxes. The [longer/shorter] a Dax is, the better it is at catching fish.

The longer versus shorter wording was the prescriptive Ideal manipulation. Immediately after hearing this, participants were shown two "Dax" images of different lengths and asked which one was better for catching fish. This was a manipulation check item that was used as an exclusion criterion.

Then, they were told "there were a bunch of Daxes in a big pile" that they would see them all one at a time and then be asked some questions about Daxes. After this, participants saw 100 Daxes of varying lengths individually. They had to click a "next" button after viewing each Dax, and the button was disabled for 500 ms after the image was changed to ensure participants had time to view them. Because the minimum viewing time was only 500 ms per item, it was possible to see all 100 items in 2 min or less, though we did not measure how long this phase of the experiment took (but we did measure total duration, see the Results section). The Daxes varied in width between 250 and 450 pixels, achieved by adding a number of pixels between 1 and 100 multiplied by 2 to an image that was originally 250 pixels in width (i.e., +2 to +200 pixels from the original width). This created 100 possible length values. Going forward, for simplicity, we will refer to length as a scale from 1 to 100, even though that does not correspond to the exact pixel values. The 100 items were a preselected normal distribution of length values with a mean length of 40 and a standard deviation of 15. Participants were given encouraging audio feedback after 25 ("Keep going, you're doing great!"), 50 ("You're halfway done! Keep it up!"), and 75 items ("You're almost done, keep going!").

After seeing the 100 items, participants were reminded of the prescriptive value of the items using the same wording as before. They then proceeded to the test item. Participants were randomly assigned to a "first-to-mind" or "random sample" condition.

In the first-to-mind condition, they were told "Think of a Dax, then use the slider to show me the Dax you thought of." There was a "start" button, and when clicked, it showed this question in text and played the audio narration, then gave a 3-s countdown, and then showed a Dax with a random length value and the circular slider in a random orientation. The slider in this item operated such that the length value (and the length of the object in the middle) changed as the slider went around the circle, from length value 1 (252 pixels) to 100 (450 pixels) and then back down to 1 as they went around. The location of 1 on the circle was given a random orientation, and the circle was not labeled, so participants could only explore the length space by moving the slider around the circle. Furthermore, before responding, they had to move the slider at least 15 degrees from its starting location, though they could return it to that location before actually submitting a response. This ensured that there was no systematic bias based on the response method (e.g., no linear scale endpoints for participants to anchor on, and no consistent starting value between participants).

In the random sample condition, the question was instead as follows:

> All the Daxes are in a machine, and when someone pushes a button, they get a Dax, but they don't get to choose which one they get. Someone pushed the button. Use the slider to show me the Dax you think they got!

This wording was chosen because a random sample from a normal frequency distribution should simply recreate the distribution, the peak of which is the average. This wording also corresponded to the random sample training item. Other than the question wording, the slider operated identically to the first-to-mind condition.

After answering the test item, they were asked an additional validation question using the slider, where they had to use the slider to show "the best Dax." This item was the same in both conditions and was included for exploratory analyses.

Participants then went on to the second learning and test phase, which was identical to the first half except that it used a different item called a "Fep" which was used for "collecting fruits and vegetables" and varied in height rather than width. The image used was a novel object created in 3D modeling software, consisting of two toruses connected to a central cube. The training phase, length values, and test phase were otherwise identical to the first half. Notably, the Ideal condition used for the second familiarization phase was the opposite of whatever was used in the first familiarization. That is, if a given participant was told that longer Daxes are better, they were told that shorter Feps are better, and vice versa. Thus, Ideal was a within-subjects manipulation (with counterbalanced order), while Question (first-to-mind vs. random sample) was a between-subjects manipulation.

Finally, after completing the second test phase (including a "best Fep" item), participants were shown a final attention check where they saw the two novel objects side by side and were asked to click on the Dax.

We recorded when the participant opened the page and when the experiment ended. Adults took an average of 705 s (~11.3 min) to complete the task ($SD = 279$ s), and children took an average of 924 s (~15.4 min; $SD = 373$ s).

Exclusion Criteria

Our only preregistered exclusion criteria in this experiment were (a) the participant had to respond to *at least one* test item (i.e., we accepted partial data) and (b) the participant had to answer the previously described two manipulation check items (one for each test block) and one attention check item correctly. Participants who got *any* of these three items wrong were excluded and replaced prior to analysis. In the end, out of 93 total exclusions (90 children and three adults), 86 were due to failing a manipulation check and seven were due to failing the final attention check.

Results

Our preregistered analysis plan started with a 2 (Question) × 4 (Age group) × 2 (Ideal) linear mixed-model analysis with a random effect of participant on intercept (because Ideal is a within-subjects factor). This analysis revealed a main effect of Ideal, $F(1, 238) = 125.25$, $p < .001$, a Question × Ideal interaction, $F(3, 238) = 21.45$, $p < .001$ and a Question × Age group interaction, $F(3, 238) = 3.14$, $p = .026$. The three-way interaction was not significant, $F(3, 238) = 1.23$, $p = .30$.

We then conducted separate paired-sample t tests of the effect of Ideal in each Question condition in each Age group, in order to test the hypothesis that participants gave different judgments to each item depending on Ideal (since Ideal was manipulated within-subjects, but the true mean of the distribution was always 40). The logic of this comparison is that if a participant did not incorporate Ideal into their judgments (i.e., the null hypothesis), then they should give the same rating for both items. See Table 1 for the results, and Figure 2 for the magnitude of the effect size for the effect of ideal in each Age group and Question condition.

Unsurprisingly, the adult results showed a substantial difference between questions. On the random sample predictions, adults showed no effect of Ideal, whereas on the first-to-mind question, adults showed a highly significant effect of Ideal. This simply replicates previous findings (Bear et al., 2020).

The key question was what would happen to children. Like adults, children of all ages showed a significant effect of Ideal on first-to-mind judgments. Importantly, however, children also showed an effect of Ideal for random sample predictions. As shown in Table 1, for 4–5-year-olds, the effect of Ideal on random sample predictions does not survive Bonferroni correction for eight tests, but for 6–7- and 8–9-year-olds, the effect is robust to correction.

Discussion

Experiment 1 successfully replicates past results with adults. Adults show a differentiation between our two questions. That is, in adults, there is an effect of the prescriptive ideal on first-to-mind judgments but no impact of the prescriptive ideal on random sample predictions. By contrast, children do not show this differentiation between the two questions. Adults and children in this experiment showed an impact of the prescriptive ideal on first-to-mind judgments, but importantly, 6–9-year-old children also showed an impact of the prescriptive ideal on random sample predictions. In general, these results suggest that children integrate prescriptive information into their representation of distributions even in situations when adults do not.

The youngest (4–5-year-old) children did not show as clear an effect of Ideal in their random sample predictions, but the exclusion rate in this age group was also extremely high (and 24 of the 42 exclusions from this age group were in the random sample condition), possibly indicating that they had difficulty understanding the task.

These results suggest support for the Undifferentiated information hypothesis, in that children's first-to-mind judgments and random sample predictions were both a combination of prescriptive and descriptive information to a roughly similar degree, suggesting that children have a single undifferentiated representation, while

Table 1
Mean Ratings in Each Condition, and (Bonferroni-Corrected) p Values of the Effect of Ideal

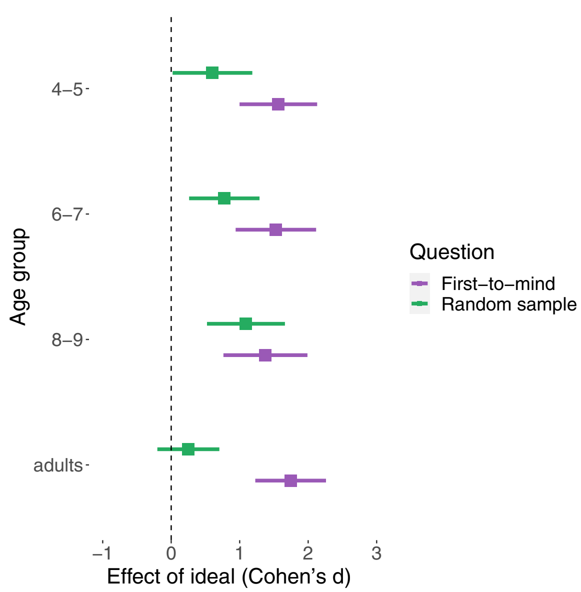
Age group	Condition	*M* high ideal (*SD*)	*M* low ideal (*SD*)	Paired *t* test *p* (Bonferroni corr.)
4–5	First-to-mind	68.31 (30.36)	22.79 (27.82)	<.001***
	Random sample	65.30 (30.01)	45.29 (36.39)	.322
6–7	First-to-mind	76.43 (30.96)	29.40 (30.58)	<.001***
	Random sample	61.09 (33.82)	34.09 (35.79)	.022*
8–9	First-to-mind	57.59 (39.92)	13.54 (20.83)	<.001***
	Random sample	65.82 (28.04)	32.90 (32.07)	.001**
Adult	First-to-mind	72.54 (31.10)	21.98 (26.71)	<.001***
	Random sample	44.23 (28.05)	38.00 (21.18)	>.5

Note. The ground truth descriptive mean was 40. Corr. = corrected.
* $p < .05$. ** $p < .01$. *** $p < .001$.

adults are able to provide random sample predictions that are unbiased by prescriptive ideal. This would also be in line with past work on how children select exemplars for categories based on distinctive traits (Foster-Hanson & Rhodes, 2019). It could also be viewed as somewhat compatible with the Prescriptive bias hypothesis, except that the effect of Ideal on children's first-to-mind judgments was not obviously greater than it was for adults' first-to-mind judgments (note the very similar effect sizes for the first-to-mind judgments in Figure 2), which is the most distinctive prediction of the prescriptive bias hypothesis.

However, there are two notable caveats regarding children's responses. First, 6–9-year-olds did not provide an unbiased sample in *any* condition, so it is unclear whether they can do so in this task. In the absence of prescriptive information, children seem to be able to track statistical information when making generalizations about categories (Rhodes & Liebenson, 2015; cf. Foster-Hanson & Lombrozo, 2022), but it is unclear whether they can do so when prescriptive information is available, and particularly when it is not aligned with the statistical regularities of a category. Second, the effect may not be driven by the prescriptive ideal per se, but rather by pure salience. That is, we emphasize the attribute associated with the "best" item, and this could bias children's responses not because it is prescriptively the best, but simply because we called attention to it. We designed Experiment 2 to examine the role of salience and to see if 6–9-year-old children can produce unbiased estimates in their random sample judgments.

Figure 2
Effect Size Magnitude (Cohen's d) for the Effect of Ideal in Each Question and Age Group Condition



Note. Error bars represent 95% confidence intervals. The effect of Ideal was clear for all first-to-mind conditions, as well as the random sample conditions for 6–7 and 8–9-year-olds (for 4–5-year-olds, the effect did not survive correction and adults showed no evidence of an effect before correction). See the online article for the color version of this figure.

Experiment 2

In this experiment, we attempted to dissociate whether (a) one endpoint of the dimension we varied in our stimuli was salient and (b) whether it was prescriptively good. To do this, we changed how we provided prescriptive information. For half the participants, they only received information about the "best" items. For the other half, they only received information about the "worst" items. We sought to test three hypotheses about these different Focus (positive vs. negative) conditions:

1. Prescriptivity hypothesis: Providing information about what is bad could allow children and adults to infer that the opposite must be good (e.g., Morris, 2003). Under this view, negatively valenced prescriptive information (i.e., only talking about the "worst" items) will be inverted in order to determine the ideal, and the ideal will influence judgments. That is, participants will use this information to infer what the "best" items are and therefore align with the "positive" condition and Experiment 1. This predicts no effects of or interactions with Focus, but an interaction between Question and Ideal in adults, and a main effect of Ideal in children.

2. Ignore negative hypothesis: Children tend to discount prescriptively bad things. Past work has suggested that up to about 6 years of age, children regard prescriptively bad actions as not only improbable but actually also impossible, and even up to age 8, they treat them as highly improbable (Shtulman & Phillips, 2018). As such, if only positively valenced prescriptive information influences first-to-mind

and (in children) random sample judgments, then negatively valenced prescriptive information may just be ignored by both children and adults (and not even be used to infer what is "good"). Under this hypothesis, there will be no effect of prescriptive information in the "negative" condition on either type of judgment at any age. This predicts an interaction between Focus and Ideal, such that the effect of Ideal is only present in the positive Focus condition.

3. Salience hypothesis: This is the primary deflationary explanation of the results of Experiment 1. Under this view, salience is the only relevant factor for children's first-to-mind and random sample predictions, and so their responses in the "negative" condition will be biased toward the "worst" item, while those in the "positive" condition will be biased toward the "best" item. This also predicts an interaction between Focus and Ideal, but in this case, it will be due to there being an effect of Ideal in both Focus conditions, but in opposite directions. The most distinctive feature of this hypothesis is that we should see a *negative* effect size for Ideal for both questions in the negative Focus conditions (i.e., judgments shifted *away from* the true prescriptive ideal).

In addition, we decided to restrict the age range of this study to 6–9-year-old children and adults, due to the exceptionally high exclusion rate of 4–5-year-olds in Experiment 1, indicating that the task might be too complex or too long for them.

Method

Transparency and openness

The adult sample and developmental sample were preregistered separately (https://osf.io/4eb5k; https://osf.io/znqmu). All materials, data, and analysis scripts can be found at https://osf.io/nrjaz/.

Participants

We preregistered (https://osf.io/4eb5k; https://osf.io/znqmu) that we would collect data from *at least* 26 participants that passed our exclusion criteria (see below) in each cell of an Age Group (6 and 7, 8 and 9, and adults) × Focus (positive vs. negative) × Question (first-to-mind vs. random sample) between-subjects design (with Ideal as a within-subjects factor), leading to a target sample size of $26 \times 3 \times 2 \times 2 = 312$ with 104 participants in each age group. As in Experiment 1, we planned to continue data collection until we reached this minimum in every cell and retain any extra data collected as a result.

A new sample of adults was recruited from Prolific, as before. Children were recruited from Children Helping Science (CHS) using the now-integrated Lookit platform to obtain parental consent and child assent (Scott & Schulz, 2017). Parents had to provide a recorded video consent, and children had to provide a recorded video assent. If either was missing, or if the child was not visible in the assent video, the data were excluded. The consent/assent videos were associated with the response by CHS's anonymous child identifier code, which was passed to the task through a query string in a redirect URL that participants were automatically sent to when they finished the video assent. After integration with Lookit, CHS now records demographics and precise ages, so we were able to record this data for Experiment 2.

Our final analysis included 109 6- and 7-year-olds (50 female, 66 male, three declined to respond), 110 8- and 9-year-olds (62 female, 54 male, two nonbinary, two declined to respond), and 109 adults (41 female, 65 male, one nonbinary, two declined to respond). The reported demographics of this sample were as follows: 54% White, 14% Asian, 6% Black or African-American, 3% Hispanic or Latino (and no other identity), 0.5% American Indian or Alaska Native, 1% "other," and 17% reported more than one race or ethnicity (and 3.7% declined to respond). An income demographic question indicated that the families who participated were mostly middle-class (and all were from the United States). An additional 71 6- and 7-year-olds (37 female, 32 male, one nonbinary, one declined to respond), 47 8- and 9-year-olds (24 female, 20 male, three declined to respond), and 56 adults (34 female, 22 male) were excluded and replaced based on preregistered exclusion criteria (see below).

Stimuli and Procedure

The stimuli and procedure were identical to Experiment 1 except for any part of the experiment that involved valence information. This affected the training and familiarization phases, as well as the secondary test item, where they were asked to use the slider to make the "best" object in Experiment 1. Full materials can be found at https://osf.io/nrjaz.

For the training, the positive condition was identical to Experiment 1, but in the negative condition, participants were instead asked the following:

> Sometimes we might want to know what the worst thing is. For example, say you took a quiz with four questions, and the teacher gave you a red X like this one for every question you got wrong. What would be the worst thing you could get on this quiz? How many Xs? Click on the picture that shows the worst number of Xs.

Similarly, the slider for this training item used red Xs rather than gold stars.

For the familiarization, any time prescriptive value was mentioned, only the positive terms were used in the positive Focus condition, and only the negative terms were used in the negative Focus condition. As a reminder, in Experiment 1, participants were told "Some [Daxes] are awful, some are very bad, some are a little bad, some are a little good, others are very good, and some are awesome." Those in the positive condition of this experiment instead heard "Some [Daxes] are a little good, others are very good, and some are awesome," while those in the negative condition heard "Some [Daxes] are a little bad, others are very bad, and some are awful." When told the relationship between the length dimension and prescriptive value, they were told, for example, "The awesome Daxes that catch the most fish are the [longest/shortest] Daxes. The [longer/shorter] a Dax is, the better it is at catching fish" in the positive condition, and "The awful Daxes that catch the fewest fish are the [shortest/longest] Daxes. The [shorter/longer] a Dax is, the worse it is at catching fish." The check item for valence was similarly modified to ask for the "worse" one in the negative condition, as was the second "best/worst" test item that followed the primary test item. The same changes were made to the "Fep" item as well.

Focus condition was manipulated entirely between subjects, so for a given participant, they would only hear about which items were "better" or "best," or they would hear about which items were "worse" or "worst." Ideal (i.e., whether the longer or shorter object was better) was still manipulated within-subjects, and for the purposes of analysis, we always examined the Ideal corresponding to the "best" option regardless of Focus condition (i.e., in the negative condition, if told the shorter ones were worse, the Ideal condition would still be "longer = better").

Adults took an average of 651 s (~10.9 min) to complete the task (SD = 277 s), and children took an average of 760 s (~12.7 min; SD = 186 s). Children were faster to complete this experiment relative to Experiment 1 most likely because they were older (no 4–5-year-olds) and the timer started after they had completed the consent, whereas in Experiment 1, the timer included the consent.

Exclusion Criteria

Our exclusion criteria consisted of two manipulation check items, one attention check, consent/assent validation, and data completeness. The attention check was identical to Experiment 1. The manipulation checks in the positive condition were identical to Experiment 1, but those in the negative condition asked which one was *worse* rather than better. The consent/assent validation is as described above. Finally, unlike Experiment 1, we elected not to accept incomplete data. Participants had to provide responses to both test items as well as the two "Best/worst" manipulation check items described above. If any of these responses were missing (due to unwillingness to respond or technical issues, which we could not distinguish), that participant was excluded from analyses and replaced. Of the total 174 exclusions (118 children and 56 adults), 33 were due to failing to respond to at least one test item, 14 children were due to an invalid consent or assent video, 19 adults were due to failing the final attention check (every child who was not excluded for another reason passed this item), and the remaining 108 (77 children and 31 adults) were excluded due to failing one of the manipulation check items. Given the strictness of these criteria in an unmoderated online testing platform, the observed exclusion rate (35%), while high for an in-lab or moderated study, was entirely within expectations. For comparison, other unmoderated developmental studies with various age groups and methods have reported exclusion rates anywhere between ~1% (Foster-Hanson et al., 2024, Study 2) and 20% (Nussenbaum et al., 2020) when failure to provide data is the only exclusion criterion, and as high as ~50% (Scott & Schulz, 2017) with stringent parental interference and participant inattention exclusion criteria. The present experiments had comprehension and inattention checks as participant-level exclusion criteria, and our exclusion rate falls predictably between the studies that used the least and most stringent criteria in past work.

Results

Our first preregistered analysis was a 3 (Age group) × 2 (Question) × 2 (Focus) × 2 (Ideal) omnibus mixed-model analysis of variance. This revealed a significant main effect of Age group, $F(2, 336) = 5.87$, $p = .003$, a significant effect of Ideal, $F(1, 336) = 42.19$, $p < .001$, a Significant Focus × Ideal interaction, $F(1, 336) = 30.61$, $p < .001$, a Significant Question × Ideal interaction, $F(1, 336) = 12.42$, $p < .001$, and a significant three-way Focus × Ideal × Age Group interaction, $F(1, 336) = 4.70$ $p = .009$. The four-way interaction was not significant, $F(2, 336) = .058$, $p = .56$. These results are compatible with all three of our hypotheses, since we observed a main effect of Ideal, a Question × Ideal interaction, and a Focus × Ideal interaction, but the critical question to distinguish between the hypotheses is whether the effect of Ideal is present in both Focus conditions.

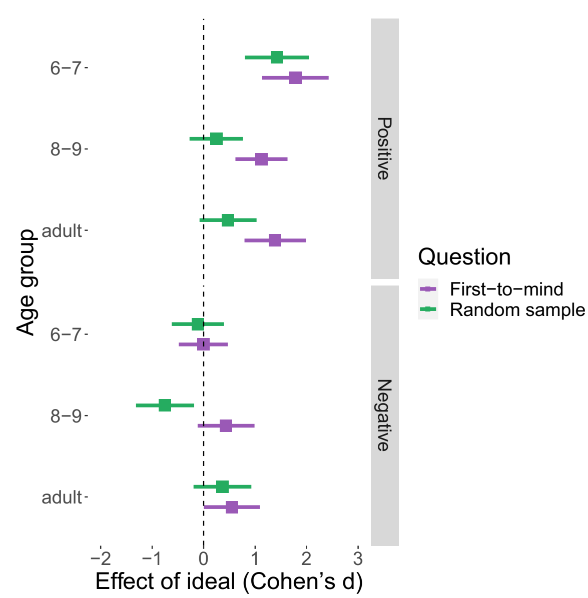
We preregistered that, regardless of the four-way interaction, we would conduct separate Age Group × Question × Ideal analyses of variance for each Focus condition. All three hypotheses predict that the positive Focus condition should largely replicate Experiment 1. There were significant main effects of Ideal, $F(1, 168) = 75.48$, $p < .001$, and Age group, $F(2, 168) = 3.53$, $p = .03$, as well as a significant Question × Ideal interaction, $F(1, 168) = 7.91$, $p = .005$, and a significant Age Group × Ideal interaction, $F(2, 168) = 5.14$, $p = .007$. In contrast to Experiment 1, there was no Age Group × Question interaction, $p = .95$, but further analyses (see below) found that 6–7-year-olds performed very similarly to Experiment 1, while 8–9-year-olds did not. Once again, we did not observe a three-way interaction, $p = .64$.

For the negative Focus condition, the salience hypothesis predicts a main effect of Ideal or a pattern similar to the positive condition but with the effects in the opposite directions (i.e., first-to-mind judgments biased toward the worst rather than best), while the prescriptivity hypothesis predicts the same pattern as the positive condition, and the ignore-negative hypothesis predicts no effect of ideal. This analysis found a main effect of age group, $F(2, 168) = 3.69$, $p = .027$, and a Question × Ideal interaction, $F(1, 168) = 4.77$, $p = .030$, but no other main effects or interactions. As there was no main effect of Ideal, the results seem to fit best with the ignore-negative hypothesis.

Figure 3 shows the effect size of Ideal in each Age group, Question, and Focus condition. To summarize a great deal of data (see Table 2), as the upper panel of the figure suggests, in the positive Focus condition, we found a similar pattern to Experiment 1: Every age group showed an effect of Ideal for first-to-mind

Figure 3

Effect Size of the Effect of Ideal in Each Age Group in Each Question Condition in Experiment 2



Note. The bottom half shows the "negative" Focus conditions, while the top half shows the "positive" Focus conditions. Error bars represent 95% confidence intervals. The "positive" conditions were similar to Experiment 1 (except the 8–9-year-olds in the random sample condition), while there were no statistically significant effects of Ideal in the "negative" conditions. See the online article for the color version of this figure.

Table 2
Mean Ratings in Each Condition in Experiment 2 and (Bonferroni-Corrected) p Values of the Effect of Ideal

Focus	Age group	Question	*M* high ideal (*SD*)	*M* low ideal (*SD*)	Paired *t* test *p* (Bonferroni corr.)
Positive	6–7	First-to-mind	77.37 (30.92)	22.81 (30.31)	<.001***
		Random sample	73.81 (29.60)	30.81 (30.78)	<.001***
	8–9	First-to-mind	65.25 (35.39)	28.81 (29.21)	<.001***
		Random sample	55.43 (32.39)	47.47 (32.81)	>.5
	Adult	First-to-mind	58.75 (28.11)	22.93 (23.18)	<.001***
		Random sample	50.67 (30.91)	36.85 (27.28)	>.5
Negative	6–7	First-to-mind	60.51 (40.32)	60.83 (40.80)	>.5
		Random sample	48.87 (28.37)	52.23 (31.25)	>.5
	8–9	First-to-mind	56.44 (37.28)	40.81 (34.45)	>.5
		Random sample	31.11 (33.03)	55.74 (32.76)	.098
	Adult	First-to-mind	53.96 (31.20)	36.50 (32.62)	>.5
		Random sample	51.85 (29.84)	41.31 (27.87)	>.5

Note. Corr. = corrected.
*** $p < .001$.

judgments, but adults showed no effect of Ideal for random sample predictions while 6–7-year-olds did. Interestingly, and contrary to the results of Experiment 1, 8–9-year-olds in this experiment did not show an effect of Ideal in their random sample predictions, suggesting that this sample of 8–9-year-olds may have been more adult-like in their responses and that this may be a potential transition period.

In contrast, the negative Focus conditions show a very different pattern. There is no effect of Ideal whatsoever for 6–7-year-olds, regardless of Question. For 8–9-year-olds, there are hints of an inverse effect (i.e., a bias toward the "worst" values) for random sample predictions, which is what would be predicted for *both* child groups and *both* questions by the salience hypothesis, but the effect does not survive Bonferroni correction for 12 comparisons. For adults, there is a slight positive effect of Ideal on first-to-mind judgments in the negative condition, which would fit the prescriptivity hypothesis, but it also does not survive correction. In short, providing exclusively negatively valenced information had no reliable effect on either first-to-mind or random sample judgments in any of the age groups studied here. While null results cannot be taken as affirmative evidence for the ignore-negative hypothesis, the prescriptivity and salience hypotheses specifically predict effects in these conditions that we did not observe, while the ignore-negative hypothesis predicts no such effects. Therefore, we conclude that our findings are more compatible with the ignore-negative hypothesis than either of the alternatives.

Our final preregistered analysis was a comparison of each condition's mean against the descriptive average of what participants were shown. We conducted two-tailed, single-sample t tests against the true mean of 40, corrected for 24 comparisons. Six to seven-year-olds in the positive Focus condition had mean ratings significantly above the true mean in both the average and first-to-mind condition when the Ideal was high, corrected $ps < .001$, while 8–9-year-olds in the positive condition had mean ratings significantly above the true mean only in the first-to-mind condition when the ideal was high, corrected $p = .003$, and adults in the positive Focus first-to-mind Question condition had a mean response significantly lower than the true average when the Ideal was low, corrected $p = .014$, and significantly higher than the true average in the corresponding high Ideal condition, corrected $p = .036$. None of the means in the negative Focus conditions deviated significantly from the true mean, corrected $ps \geq .12$. Thus, even 6–7-year-old children were able to produce an unbiased random sample prediction in the negative Focus conditions.

Discussion

This second experiment was designed to determine whether the results obtained in Experiment 1 genuinely reflected an impact of the prescriptive ideal or whether they simply reflected the impact of a salience confound. When the value that was made salient was the same as the prescriptive ideal, we again found that 6–7-year-old children's random sample predictions were shifted toward that value. By contrast, when the value that was made salient was the opposite of the prescriptive ideal, children's random sample predictions were not shifted toward that value. Instead, their predictions were, in aggregate, an unbiased estimate of the true sample mean. This result indicates that the effect is not driven entirely by salience and does indeed reflect an impact of the prescriptive ideal.

General Discussion

We started with broad questions about how people use prescriptive and descriptive information for different tasks, and how the use of these different kinds of information might change over development. Experiment 1 found that children's first-to-mind judgments are very similar to adults, in that they are a combination of the peak of a descriptive frequency distribution and a prescriptive ideal. However, children's predictions of random samples are nearly identical to their "first-to-mind" judgments and are influenced by prescriptive information in the same way. Experiment 2 ruled out a salience explanation for this result and further found that only positive prescriptive information influences both children's and adults' first-to-mind judgments.

The results of Experiment 1 and the positive Focus condition of Experiment 2 are clearly most compatible with the "undifferentiated information" developmental hypothesis (Hypothesis 3). That is, children incorporate both descriptive and prescriptive information into their internal representations. However, unlike adults, they cannot consistently separate them when asked to make a prediction that should normatively ignore prescriptive value (though 8–9-year-olds were able to provide unbiased random sample predictions in

Experiment 2). This conclusion provides a clear perspective on earlier work that showed children's difficulties in separating prescriptive and descriptive information (Barsalou, 1985; Kim & Murphy, 2011; Shtulman & Phillips, 2018; Tisak & Turiel, 1988): It is not that they ignore either type of information, they just treat both of them as informative about both what "should" happen, and about what will happen in the future. Indeed, we would explain these previous results as indicating that younger children's judgments of likelihood, permissibility, and typicality all draw from this shared underlying distribution.[2]

Experiment 2 provides some deeper insight into some of the underlying mechanisms that drive these first-to-mind judgments. In the models used by past work (Bear & Knobe, 2017; Bear et al., 2020), the full distribution of prescriptive value is used as a multiplier or incorporated into a softmax model as an exponent, regardless of whether it is positive or negative. In mechanistic terms that represents something like the internal representation of the frequency distribution being weighted by prescriptive value, such that high-value items are oversampled and low-value items are undersampled. However, our results suggest that only positive prescriptive value has an influence on first-to-mind judgments, while negative prescriptive value is simply ignored. Indeed, in some of these past experiments, it is unclear whether negative prescriptive information was provided. In the case of Bear et al. (2020)'s Experiments 2 and 3, they presented "grades" to represent quality that ranged from D to A, but described the general prescriptive principles only in positive terms (e.g., "longer is better"). As such, the existing models are arguably restricted to positive prescriptive information implicitly, but our results suggest that in cases where both positive and negative prescriptive information are available, models should be explicitly restricted to positive prescriptive information.

One concern readers might have is that younger children did not track frequency information at all. Under this view, in Experiment 1 and the positive Focus condition of Experiment 2, children's answers to both questions were just based on prescriptive information, and their responses were completely random in the Negative focus condition (since the center of the scale [50] is close to the peak of the frequency distribution [40] it would be hard to identify uniform randomness by means alone). However, recall that in each experiment, we also asked participants to use the slider to show us the "best" (or in the negative Focus condition of Experiment 2, "worst") Dax or Fep after answering the primary question, as an exploratory dependent variable. In past work on category representation, even though children seemed to pick a mix of descriptive and prescriptive information to use as an exemplar of a category, they were nonetheless able to identify the "best" without difficulty (Foster-Hanson & Rhodes, 2019). Indeed, identifying the "best" or "worst" may be based on a different kind of process than first-to-mind judgments or random sample predictions: Children in this age range have a tendency to interpret superlatives in an absolute or categorical way (Tieu & Shen, 2015), and a categorical interpretation does not require sampling from a distribution at all, just finding the most extreme value that the slider can produce.

These "best" (or "worst") ratings, which we report in the Supplemental Tables S1 and S2, tell us what children's judgments look like when they are based on prescriptive information alone, and they are markedly different from their first-to-mind and random sample responses. In particular, "best" (or "worst") ratings are always at least 15 points further from the true mean than the corresponding first-to-mind or random sample response. This strongly indicates that children's first-to-mind judgments are, in fact, a combination of descriptive and prescriptive information; children would have to be tracking frequency information in some way. In addition, in the negative Focus condition of Experiment 2, children provided similarly extreme responses when asked for the "worst" item (means between 2 and 15 for the high ideal, and 85 and 92 for the low ideal), despite Ideal having no detectable effect on our primary dependent variables in these conditions. This suggests that children did pay attention to the negative prescriptive information, but it was not incorporated into the underlying distribution they used to make their random sample predictions and first-to-mind judgments.

Limitations and Puzzles

Our conclusions are constrained by a number of limitations of our methods, and there are some patterns in the results for which we have no definitive explanation. One limitation is that, for confidentiality and privacy reasons, we did not video- or audio-record participants as they completed the task. For children in particular, this means that we do not know how much assistance they received from their parents. If the parents were providing extensive guidance or simply doing the task themselves, we would have expected much more similarity between the adult and child groups, so we are confident that our data do, in fact, come from children. However, if we want to better understand *how* children complete this task, future studies will benefit from either being conducted in person or with recording.

Another limitation of the methods is that our demographic information is sparse, and for Experiment 1, basically nonexistent. There has been some debate about whether we should take the demographics of online samples as representative even within a given culture or country (Lourenco & Tasimi, 2020; Sheskin et al., 2020), which certainly constrains how much we should generalize our conclusions. All of the work on these first-to-mind judgments to date (that we know of) has focused on English-speaking participants mostly from North America. While our results suggest a sort of default optimism (i.e., we tend to think of things that are both likely and good), some past work has suggested that optimism is a trait that varies systematically between cultures (Chang, 1996). The tendency to ignore negative prescriptive information, and the tendency to incorporate positive prescriptive information into first-to-mind judgments in adults and random sample predictions in 6–7-year-old children, may be absent or show opposite patterns in other cultures.

However, more specific to this project, we recruited from different populations in Experiments 1 and 2. The version of Children Helping Science that existed when we ran Experiment 1 was effectively just a central hub for posting links to studies on other websites. It had no set user base, and studies were advertised via Facebook and other social media systems, through existing lab databases, and more. By the time we conducted Experiment 2, CHS had merged with Lookit, and we used the Lookit functionality for our consent process. This gave us much more demographic information, but also meant that our participants were all people who were part of the Lookit database, who had signed up to be contacted for multiple studies, and who were likely to be engaged participants.

[2] It is worth noting, however, that recent work has suggested that judgments of whether something is categorically possible or not may operate under different principles (Cesana-Arlotti et al., 2025).

The difference in population between the two experiments is particularly relevant when considering one of the puzzles in our results: that 8–9-year-olds in Experiment 2's "positive" focus condition did not show a significant effect of Ideal on random sample judgments, while in Experiment 1, they showed a very clear effect. There are three plausible, and not mutually exclusive, explanations for this difference. One is that it is in fact a result of the difference in recruitment methods, but since we were unable to get detailed demographics for the participants in Experiment 1, it is hard to formulate hypotheses about what specific differences might impact this task (e.g., experience participating in research studies, household income, parental education, etc.).

A second possibility is that 8–9 is a transitional point in development where children begin to be more adult-like in how they predict the outcome of random samples, and our first sample fell more on one side of that transition while the second fell on the other side, either due to the difference in recruitment or just by chance. Finally, a third possibility is that the slight change to the wording in the positive Focus condition in Experiment 2 (i.e., not mentioning the "worst" members of a category at all) may have changed how children in this age range approached the random sample predictions. It is unclear why this would only matter for 8–9-year-olds but not 6–7-year-olds, but we acknowledge that it is a possibility. Future work that attempts to replicate these results could examine these possibilities directly, but with the available data, we can only speculate.

Another oddity is that 8–9-year-olds showed a hint of an effect of Ideal on random sample predictions in the negative Focus condition in Experiment 2, though the effect did not survive correction for multiple comparisons. Since it does not survive correction, we believe this is simply a spurious effect, but it might fit with a developmental trend in the literature, related to the point about "optimism" raised above in reference to possible cross-cultural differences. In particular, there is a decline in optimism when making predictions about the future between the ages of 6–7 and 8–9 (Leonard & Sommerville, 2025). Extending this idea to the current experiment, the random sample predictions by the 8–9-year-old group could have been biased toward negative prescriptive valence in both Focus conditions (a sort of pessimistic bias), resulting in the null effect in the positive condition and the hint of a negative-bias effect in the negative condition. In this account, first-to-mind judgments, which are not predictions about the future, would not be affected. However, we would only give this account serious consideration if future replications found a more consistent negative valence bias in similar prediction tasks with this age group.

What Changes Over Development?

Our results align with well-documented effects in the developmental literature. The notion that even very young children sample responses from memory is well established in domains exploring children's causal reasoning (e.g., see Bonawitz et al., 2014 for a review). However, our findings extend this body of work by demonstrating that children's first-to-mind samples are biased by positive prescriptive information (much like adults'). Furthermore, our results suggest that not only does positive prescriptive information influence recall processes in first-to-mind judgments as it does in adulthood but also prescriptive information may also lead 6- and 7-year-old children to general recall errors when predicting random samples. There is evidence that other kinds of information bias children's memory for descriptive information. For example, previous research has shown that preschool and early elementary-age children's memory tends to be biased toward category prototypes (e.g., Duffy et al., 2006; Foster-Hanson et al., 2024; Persaud et al., 2021), suggesting that, like adults, children rely on strategies such as regression to the category mean to compensate for noisy episodic traces. That positive prescriptive information leads to a similar effect suggests it is playing a role similar to "prototypes" in other nonprescriptive domains, though the category representation literature suggests that prototypes may themselves be combinations of descriptive and prescriptive information (e.g., Foster-Hanson & Rhodes, 2019). Indeed, prototypes that are idealized can shift children's expectations about the underlying distribution of prescriptively good qualities in a population (Foster-Hanson et al., 2024).

Adults, on the other hand, were able to produce relatively unbiased predictions about random samples, as were 8–9-year-old children in Experiment 2. The question then is: what factors are changing over development to facilitate the separation of descriptive information from prescriptive information in adulthood? One possibility is that adults (and perhaps older children) are able to deploy metacognitive strategies that allow them to "filter" the prescriptive information out of predictions that it normatively should not influence. Children may differ from adults because they are either unable to inhibit prescriptive information when predicting random samples (due to computational complexity or inhibitory control), or the way that they interpret the random sample question does not lead them to think that they need to ignore prescriptive information, that is, a normative way of "random" sampling to a child involves incorporating both what does happen and what *should*. Notably, if, as we suggest above, 8–9 years of age is a transition period from a more child-like to more adult-like response pattern, it may be due to the emergence of more sophisticated memory searches characterized by increased strategic retrieval, better use of semantic associations, and greater cognitive control (e.g., Paz-Alonso et al., 2009), memory regulation strategies (e.g., Ghetti et al., 2010), and explicit metacognitive skills involved in encoding (e.g., DeMarie & Ferron, 2003). These findings point to an exciting direction for future research at the intersection of these developing abilities.

Conclusion

Our results find that, like adults, children represent both descriptive and positive prescriptive information about the world and use both together to generate the "first thing that comes to mind." At the same time, there seems to be a developmental shift in the ability to filter out positive prescriptive information for tasks where a purely descriptive representation is normatively preferable (at least to adults). In terms of our understanding of the underlying computations, however, our results have left many open questions. There is much that we have yet to understand about how children (and adults) use their experience and their beliefs about prescriptive norms to predict events in the world and what downstream influences those predictions might have on their behavior and decision making. Of course, our first guess that comes to mind about how this all works is likely to be biased.

References

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

Bass, I., Bonawitz, E., Hawthorne-Madell, D., Vong, W. K., Goodman, N. D., & Gweon, H. (2022). The effects of information utility and teachers' knowledge on evaluations of under-informative pedagogy across development. *Cognition*, *222*, Article 104999. https://doi.org/10.1016/j.cognition.2021.104999

Bass, I., Gopnik, A., Hanson, M., Ramarajan, D., Shafto, P., Wellman, H., & Bonawitz, E. (2019). Children's developing theory of mind and pedagogical evidence selection. *Developmental Psychology*, *55*(2), 286–302. https://doi.org/10.1037/dev0000642

Bass, I., Mahaffey, E., & Bonawitz, E. (2023). Children use teachers' beliefs about their abilities to calibrate explore–exploit decisions. *Topics in Cognitive Science*. Advance online publication. https://doi.org/10.1111/tops.12714

Bear, A., Bensinger, S., Jara-Ettinger, J., Knobe, J., & Cushman, F. (2020). What comes to mind? *Cognition*, *194*, Article 104057. https://doi.org/10.1016/j.cognition.2019.104057

Bear, A., & Knobe, J. (2017). Normality: Part descriptive, part prescriptive. *Cognition*, *167*, 25–37. https://doi.org/10.1016/j.cognition.2016.10.024

Bonawitz, E. B., Denison, S., Griffiths, T. L., & Gopnik, A. (2014). Probabilistic models, learning algorithms, and response variability: Sampling in cognitive development. *Trends in Cognitive Sciences*, *18*(10), 497–500. https://doi.org/10.1016/j.tics.2014.06.006

Bonawitz, E. B., & Lombrozo, T. (2012). Occam's rattle: Children's use of simplicity and probability to constrain inference. *Developmental Psychology*, *48*(4), 1156–1164. https://doi.org/10.1037/a0026471

Cesana-Arlotti, N., Jáuregui, S., Mazalik, P., Nichols, S., & Halberda, J. (2025). *Of impossible cupcakes and necessary peppers: Preschoolers deploy logical concepts to make sound modal decisions* [Paper presentation]. Presentation, 9 January 2025, Budapest CEU Conference on Cognitive Development, Budapest, Hungary.

Chang, E. C. (1996). Cultural differences in optimism, pessimism, and coping: Predictors of subsequent adjustment in Asian American and Caucasian American college students. *Journal of Counseling Psychology*, *43*(1), 113–123. https://doi.org/10.1037/0022-0167.43.1.113

DeMarie, D., & Ferron, J. (2003). Capacity, strategies, and metamemory: Tests of a three-factor model of memory development. *Journal of Experimental Child Psychology*, *84*(3), 167–193. https://doi.org/10.1016/S0022-0965(03)00004-3

Denison, S., Bonawitz, E., Gopnik, A., & Griffiths, T. L. (2013). Rational variability in children's causal inferences: The sampling hypothesis. *Cognition*, *126*(2), 285–300. https://doi.org/10.1016/j.cognition.2012.10.010

Duffy, S., Huttenlocher, J., & Crawford, L. E. (2006). Children use categories to maximize accuracy in estimation. *Developmental Science*, *9*(6), 597–603. https://doi.org/10.1111/j.1467-7687.2006.00538.x

Foster-Hanson, E., & Lombrozo, T. (2022). How "is" shapes "ought" for folk-biological concepts. *Cognitive Psychology*, *139*, Article 101507. https://doi.org/10.1016/j.cogpsych.2022.101507

Foster-Hanson, E., & Rhodes, M. (2019). Is the most representative skunk the average or the stinkiest? Developmental changes in representations of biological categories. *Cognitive Psychology*, *110*, 1–15. https://doi.org/10.1016/j.cogpsych.2018.12.004

Foster-Hanson, E., Roberts, S. O., Gelman, S. A., & Rhodes, M. (2021). Categories convey prescriptive information across domains and development. *Journal of Experimental Child Psychology*, *212*, Article 105231. https://doi.org/10.1016/j.jecp.2021.105231

Foster-Hanson, E., Ziska, K. M., & Rhodes, M. (2024). How cultural input shapes the development of idealized biological prototypes. *Journal of Cognition and Development*. Advance online publication. https://doi.org/10.1080/15248372.2024.2409680

Ghetti, S., Castelli, P., & Lyons, K. E. (2010). Knowing about not remembering: Developmental dissociations in lack-of-memory monitoring. *Developmental Science*, *13*(4), 611–621. https://doi.org/10.1111/j.1467-7687.2009.00908.x

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

Leonard, J. A., & Sommerville, J. A. (2025). A unified account of why optimism declines in childhood. *Nature Reviews Psychology*, *4*(1), 35–48. https://doi.org/10.1038/s44159-024-00384-z

Lourenco, S. F., & Tasimi, A. (2020). No participant left behind: Conducting science during covid-19. *Trends in Cognitive Sciences*, *24*(8), 583–584. https://doi.org/10.1016/j.tics.2020.05.003

Morris, B. J. (2003). Opposites attract: The role of predicate dimensionality in preschool children's processing of negations. *Journal of Child Language*, *30*(2), 419–440. https://doi.org/10.1017/S0305000903005610

Nussenbaum, K., Scheuplein, M., Phaneuf, C. V., Evans, M. D., & Hartley, C. A. (2020). Moving developmental research online: Comparing in-lab and web-based studies of model-based reinforcement learning. *Collabra: Psychology*, *6*(1), Article 17213. https://doi.org/10.1525/collabra.17213

Paz-Alonso, P. M., Ghetti, S., Matlen, B. J., Anderson, M. C., & Bunge, S. A. (2009). Memory suppression is an active process that improves over childhood. *Frontiers in Human Neuroscience*, *3*, Article 24. https://doi.org/10.3389/neuro.09.024.2009

Persaud, K., Macias, C., Hemmer, P., & Bonawitz, E. (2021). Evaluating recall error in preschoolers: Category expectations influence episodic memory for color. *Cognitive Psychology*, *124*, Article 101357. https://doi.org/10.1016/j.cogpsych.2020.101357

Phillips, J., Morris, A., & Cushman, F. (2019). How we know what not to think. *Trends in Cognitive Sciences*, *23*(12), 1026–1040. https://doi.org/10.1016/j.tics.2019.09.007

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*, Article 867. https://doi.org/10.3389/fpsyg.2015.00867

Rhodes, M., & Liebenson, 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.cogpsych.2015.07.003

Roberts, S. O., Gelman, S. A., & Ho, A. K. (2017). So it is, so it shall be: Group regularities license children's prescriptive judgments. *Cognitive Science*, *41*(Suppl. 3), 576–600. https://doi.org/10.1111/cogs.12443

Scott, K., & Schulz, L. (2017). Lookit (part 1): A new online platform for developmental research. *Open Mind: Discoveries in Cognitive Science*, *1*(1), 4–14. https://doi.org/10.1162/OPMI_a_00002

Sheskin, M., Scott, K., Mills, C. M., Bergelson, E., Bonawitz, E., Spelke, E. S., Fei-Fei, L., Keil, F. C., Gweon, H., Tenenbaum, J. B., Jara-Ettinger, J., Adolph, K. E., Rhodes, M., Frank, M. C., Mehr, S. A., & Schulz, L. (2020). Online developmental science to foster innovation, access, and impact. *Trends in Cognitive Sciences*, *24*(9), 675–678. https://doi.org/10.1016/j.tics.2020.06.004

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., & 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

Téglás, E., & Bonatti, L. L. (2016). Infants anticipate probabilistic but not deterministic outcomes. *Cognition*, *157*, 227–236. https://doi.org/10.1016/j.cognition.2016.09.003

Tieu, L., & Shen, Z. (2015). Searching for absolute and relative readings of superlatives: A second experiment. *BUCLD 39: Proceedings of the 39th Annual Boston University Conference on Language Development, November 7–9, 2014, Boston University* (pp. 424–436).

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

Tworek, C. M., & Cimpian, A. (2016). Why do people tend to infer "ought" from "is"? The role of biases in explanation. *Psychological Science*, *27*(8), 1109–1122. https://doi.org/10.1177/0956797616650875

Xu, F., & Denison, S. (2009). Statistical inference and sensitivity to sampling in 11-month-old infants. *Cognition*, *112*(1), 97–104. https://doi.org/10.1016/j.cognition.2009.04.006

Xu, F., & Garcia, V. (2008). Intuitive statistics by 8-month-old infants. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(13), 5012–5015. https://doi.org/10.1073/pnas.0704450105

Received October 29, 2024
Revision received April 8, 2025
Accepted April 9, 2025 ■