This postprint was originally published by the APA as:
Jones, A., Markant, D. B., Pachur, T., Gopnik, A., & Ruggeri, A. (2021). **How is the hypothesis space represented? Evidence from young children's active search and predictions in a multiple-cue inference task.** *Developmental Psychology*, *57*(7), 1080–1093.
https://doi.org/10.1037/dev0001201

Supplementary material to this article is available. For more information see
http://hdl.handle.net/21.11116/0000-0007-1E82-5

This paper was published by the American Psychological Association as:
Jones, A., Markant, D. B., Pachur, T., Gopnik, A., & Ruggeri, A. (2021). How is the hypothesis space represented? Evidence from young children's active search and predictions in a multiple-cue inference task. *Developmental Psychology*, *57*(7), 1080–1093.
https://doi.org/10.1037/dev0001201

How Is the Hypothesis Space Represented? Evidence From Young Children's Active Search and Predictions in a Multiple-Cue Inference Task

Angela Jones, Douglas B. Markant, Thorsten Pachur, Alison Gopnik, Azzurra Ruggeri

Author Affiliations:
Angela Jones: MPRG iSearch — Information Search, Ecological and Active Learning Research with Children, Max Planck Institute for Human Development, Berlin, Germany. School of Education, Technical University of Munich.
Douglas B. Markant: Department of Psychological Science, University of North Carolina, Charlotte.
Thorsten Pachur: Center for Adaptive Rationality, Max Planck Institute for Human Development, Berlin, Germany
Alison Gopnik: Department of Psychology, University of California, Berkeley
Azzurra Ruggeri: MPRG iSearch — Information Search, Ecological and Active Learning Research with Children, Max Planck Institute for Human Development, Berlin, Germany. School of Education, Technical University of Munich. Email: ruggeri@mpib-berlin.mpg.de.

Abstract

To successfully navigate an uncertain world, one has to learn the relationship between cues (e.g., wind speed, atmospheric pressure) and outcomes (e.g., rain). When learning, it is possible to actively manipulate the cue values to test hypotheses about this relationship directly. Across two studies, we investigated how 5- to 7-year-olds actively learned cue-outcome relationships, and what their behavior revealed about how they represented the hypothesis space. Children learned how two cues (color and shape) predicted some monsters' relative speed, by selecting which monster pairs to see racing. We compared two computational models in their ability to capture children's behavior: a *cue-abstraction* model, which organizes the hypothesis space based on abstracted cue-outcome relationships, and a *permutation-based* model, which represents the hypothesis space based on the relative speed of individual monsters. The results of Study 1 (26 five-year-olds, 14 female and 25 six-year-olds, 15 female; predominantly White, fluent in English) provided the first evidence that 5- and 6-year-olds can use cue-abstraction hypothesis space representations when provided with scaffolding. However, Study 2 (65 five-year-olds, 33 female; 67 six-year-olds, 33 female; 68 seven-year-olds, 33 female; predominantly White, fluent in German) showed that young children were best described by the permutation-based model, and that only 7-year-olds, when provided with memory aids, were best captured by the cue-abstraction model. Overall, our results highlight the guiding role of the hypothesis space for active search and learning, suggesting that these two phases might trigger different representations, and indicating a developmental shift in how children represent the hypothesis space.

Keywords: multiple-cue inference, active learning, hypothesis space, representation, children

Supplementary material to this article is available. For more information see http://hdl.handle.net/21.11116/0000-0007-1E82-5).

"Stick out your tongue and say *aaaah*. Have you been feeling tired or stressed lately? Any headaches?" Doctors usually ask many such targeted questions about patients' symptoms and run additional tests to make a diagnosis. Knowing which questions to ask or tests to run, and being able to interpret symptoms, answers, and results to identify what illness a patient is most likely to suffer from is a highly specialized skill. However, it is also an ability that people rely on in much more mundane situations. Imagine you are attending your first school's sack race. Because you are the youngest racer, you get to choose whether to race against Caren or Jason. You have seen a couple of races already: Mary, a very tall girl, won against Bob, and Mike, also pretty tall, won against Lucy. One way to represent what you have learned is through a rank ordering of individuals: So far you have learned that Mary is faster than Bob, and that Mike is faster than Lucy. In this case, you would still have no idea which opponent you are more likely to

The authors have no conflict of interest to declare. We thank Costanza De Simone, Theresa Niemann, Sarah Vega, Ari Pascarella, and Patricia Sueya Li for assistance in recruiting and data collection. We would also like to thank Federico Meini for the software development and technical support. We thank all the children from the Laurence Hall of Science in Berkeley, the Madera Primary School in El Cerrito, and the Natural History Museum and Labyrinth Museum in Berlin for providing research space. Please note that we have referenced Open Science Framework (OSF) materials in this article. The video illustrating the procedure of Study 2, the scripts for both studies and the pseudonymized data can be found at this OSF link (https://osf.io/d75bm/?view_only=ab671227125043d1b7a4d7d0cedddd61).
Correspondence concerning this article should be addressed to Azzurra Ruggeri, MPRG iSearch — Information Search, Ecological and Active Learning Research with Children, Max Planck Institute for Human Development, Lentzeallee 94, 14195 Berlin, Germany. Email: ruggeri@mpib-berlin.mpg.de

win against, and observing 10 or 20 more races would not help much—unless you could see Caren and Jason competing against each other, of course. However, you would have gathered more information from those two races you observed if you had tried to understand *what makes people faster* at this game. Why did Mary and Mike win? Maybe because they are so tall, or because they had a thinner sack? Organizing your hypotheses in this way, that is, identifying salient characteristics which may be related to the outcome, has implications for how one should search. In this example, the most information can be gathered by watching a tall child with a heavy sack race against a short child with a thinner sack. This information might help you to choose between Caren, who is shorter than you but has an ultra-professional aerodynamic sack, or Jason, who is much taller than you but has a thick black leather sack.

This example illustrates how the way we represent the set of hypotheses under consideration—by organizing them in hierarchical groups according to more abstract shared features (i.e., using a *cue-abstraction* representation) or by only considering the individual rankings (i.e., a *permutation-based* representation)—guides our information search and impacts our ability to predict the outcome. In turn, search behavior and predictions may reveal underlying hypothesis space representations, as shown in adults (Juslin et al., 2003; Markant & Gureckis, 2014; Markant et al., 2016). Juslin, Olsson, and Olsson (2003) demonstrated that adults rely on either cue-abstraction-based or exemplar-based strategies when solving multiple-cue inference tasks (e.g., learning to predict whether a subspecies of frog is toxic based on four cues). Cue abstraction involves reasoning about the abstracted cue-outcome relationships, and is a highly efficient approach because it enables rapid learning and generalization (Einhorn et al., 1979; Juslin et al., 2003; Juslin, Olsson, & Olsson, 2003). In exemplar-based reasoning, holistic memories for experienced items are stored with their associated criterion values. In this article, we consider a form of exemplar-based reasoning in which learners consider possible rankings of previously seen exemplars, which we refer to as a permutation-based hypothesis space.

However, it is still unclear what type of hypothesis space representations children use. Von Helversen et al. (2010) investigated this question from a developmental perspective, and found that while the majority of adults were best described by a cue-abstraction strategy (see also Enkvist et al., 2006; Juslin et al., 2003; Juslin, Olsson, & Olsson, 2003; Trippas & Pachur, 2019), 9- to 11-year-old children were more likely to rely on similarity-based processes, that is, they made inferences based on how much a new object resembled previously encountered ones. Reliance on cue abstraction was associated with better performance for adults but not for children, suggesting that even those children who approached the task with the more efficient strategy struggled to implement it correctly (cf. Pachur & Olsson, 2012). However, we note that children's performance was generally low, suggesting that this task may have been too difficult for them and may not have captured the full scope of children's abilities. In addition, the training phase of the task was not self-directed and participants instead made judgments for a predetermined sequence of objects, which limits insights into the hypotheses they entertained.

Like adults, older children may be able to flexibly rely on different representations depending on the demands of the task, while younger children may be constrained by the development of cognitive factors such as working memory. Cue abstraction may be especially sensitive to developmental shifts in attentional control and working memory, as it requires maintaining and reasoning about multiple, overlapping layers of cue-outcome relations. Indeed, there is evidence that cue abstraction induces higher cognitive load than exemplar-based reasoning (referred to as similarity-based reasoning; Hoffmann et al., 2013), possibly due to the additional recruitment of executive functions and working memory compared with exemplar-based models (Smith et al., 1998). On the other hand, it could also be argued that reasoning about the relative ranks of exemplars (as in the sack race example above) may incur higher costs because there are a large number of possible hypotheses (i.e., different permutations) that need to be considered.

In this article, we explored how 5- to 7-year-olds searched for information that would enable them to make accurate predictions about an outcome (the winner of a race) based on multiple cues (e.g., height and weight of the sack). Specifically, we implemented, to our knowledge for the first time, a child-friendly *active learning* paradigm in which young children could search for information by selecting the cue configurations for which they wanted to observe the outcome. Allowing learners to control the selection of information during learning is more likely to provide richer insights into the hypotheses they entertain while they learn, compared with the passive paradigms implemented in previous work. Moreover, using computational modeling, we investigated what children's behavior revealed about their representation of the hypothesis space during both learning and prediction.

The Development of Information Search Across Childhood

The ability to search for information emerges at a very early age (Cook et al., 2011; Legare et al., 2013; McCormack et al., 2016; Ruggeri et al., 2019). Preschoolers are more likely to explore when they are presented with confounded evidence—that is, when they are uncertain about the causal mechanism at work (Cook et al., 2011; L. E. Schulz & Bonawitz, 2007)—or when they face evidence that violates their prior beliefs (e.g., Bonawitz et al., 2012) and infants already prefer to explore surprising events (Sim & Xu, 2017; Stahl & Feigenson, 2015). However, the ability to search for information *efficiently* is subject to considerable developmental changes. For example, despite being able to select the more informative of two given questions already at age 5 (Ruggeri et al., 2017), children do not start to implement effective question-asking strategies consistently until age 10 (Herwig, 1982; Mosher & Hornsby, 1966; Ruggeri et al., 2016; Ruggeri & Feufel, 2015; Ruggeri & Lombrozo, 2015). This is also supported by process-tracing studies that examined children's information search using information boards, where participants have to look up information about different cues for a set of options (e.g., for a set of bikes, the price, number of gears, and color) to make a decision (e.g., which bike to buy). These studies show consistent developmental improvements in search efficiency between the ages of 7 and 14, with younger children searching more exhaustively and in a less systematic manner than older children (Betsch et al., 2014, 2016; Davidson, 1991a, 1991b; Gregan-Paxton & John, 1995, 1997; Howse et al., 2003).

It is currently unclear, however, what drives the observed developmental differences in search efficiency. Previous studies suggest that older children are more efficient than younger children because they are more systematic—in the sense that they are better able to focus on the dimensions that are more important for making the decision (Betsch et al., 2016; Davidson, 1991b; Horn et al., 2016; Mata et al., 2011; Ruggeri & Katsikopoulos, 2013). Recent research also suggests that one crucial source of developmental change in information search efficiency lies in children's stopping rules: Children are considerably more likely than adults to continue their search for information beyond the point at which a decision can be made (e.g., when a single hypothesis remains; Ruggeri et al., 2016). Another proposed explanation for young children's limited efficiency in information search is that they have difficulty going beyond the object level, that is, they fail to spontaneously identify, represent and reason with more abstract task structures. Consistent with this idea, Ruggeri and colleagues (Ruggeri et al., 2021; Ruggeri & Feufel, 2015) found that scaffolding more abstract representations of the hypotheses in the 20-questions game helped 4- to 10-year-olds to ask more informative questions. For example, Ruggeri and Feufel (2015) presented 7- and 10-year-old children and adults with 20 cards, each presenting a word label (e.g., "dog" or "sheep"). Participants were randomly assigned to one of two experimental conditions that differed in the level of the abstraction of the label: a basic-level condition (e.g., "dog") or a subordinate-level condition (e.g., "dalmatian"). Participants were more likely to ask effective questions, that is, ones that targeted the objects' categories (e.g., "Is it a pet?") rather than individual objects, in the former condition than in the latter. This suggests that providing more abstract labels facilitated a shift away from reasoning based on individual objects when generating questions. The study showed that the ability to generate more abstract features for given objects (e.g., "a dog is a mammal") also improves between age 7 and age 10 (see also Herwig, 1982). In summary, developmental differences in the ability to represent task-relevant information in an abstract way may drive, or at least strongly contribute to, developmental differences in *information search*.

Overview of the Studies

In this article, we report two experimental studies that investigate the early emergence of the ability to actively learn cue-outcome relationships and to make accurate predictions about new objects. The studies involve an active learning paradigm in which 5- to 7-year-old children were presented with four monsters (see top left panel in Figure 1)and tasked to find out which kinds of monsters were faster (for a similar paradigm in research with adults, see Parpart et al., 2015). In an active learning phase, they could select which monster pairs to see running in a race to learn how two cues (color and shape) predicted the monsters' relative speed. In a subsequent test phase, the children were asked to predict the winner of races between novel monsters and to construct a "podium" in which they ranked the four monsters seen during learning.

We modeled children's behavior in the search phase assuming that they would choose to observe the race they thought would be the most informative, that is, with the highest *expected information gain* (EIG), at each step of the search. EIG quantifies the usefulness of a query based on how much new information it is expected to provide, that is, how many hypotheses will be ruled out after seeing the outcome of that query, and with what likelihood. Formally, EIG expresses the reduction of entropy, or the uncertainty as to which hypothesis is correct, upon making a query and observing its outcome (Shannon, 1948). Note that maximizing EIG at every step of the search is a greedy, myopic learning strategy and is not necessarily always the most effective long-term approach (Meder et al., 2019). Crucially, which query grants the maximum EIG depends on how the hypothesis space is structured, and different representations of the hypothesis space predict different queries.

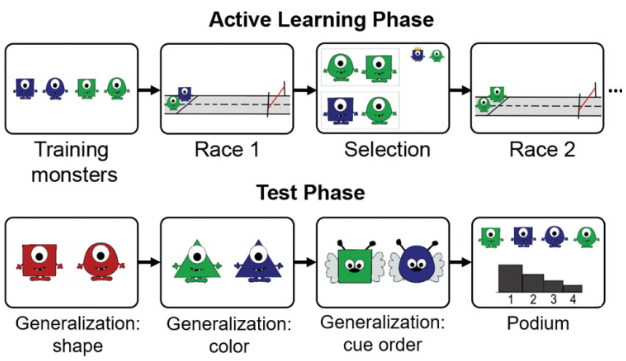

Figure 1. *Experimental Procedure in Study 1*

In the current paradigm, representing the hypothesis space in terms of abstracted cue-outcome relationships requires encoding both the cue direction and order (i.e., which color is faster, which shape is faster, and whether color or shape is more important for predicting a monster's speed; see Table 1).

Abstracting the relationship between the cues and the outcome from the pairs of monsters encountered might be rather challenging for children, and in particular for younger children, who have been shown to struggle with such abstractions (Herwig, 1982; Ruggeri et al., 2021; Ruggeri & Feufel, 2015). Abstracting the cue order (i.e., whether color or shape was more important for determining relative speed), a second-order cue, may be particularly difficult for them.

The goal of Study 1 was to explore to what extent young children are already able to represent the hypothesis space in terms of cue abstraction (see Table 1). We tested 5- and 6-year-olds' ability to actively select cue configurations, and to subsequently make inferences that are captured by a cue-abstraction representation of the hypothesis space. We designed this task to make it as easy as possible for children to adopt a cue-abstraction representation of the hypothesis space, implementing a forced-choice paradigm, making sure they collected all the information needed to make cue-abstraction-based judgments and providing memory aids. Nonetheless, we expected that 5-year-olds would be less likely than 6-year-olds to make the most informative selections and accurate generalizations as predicted by a cue-abstraction model. After establishing that children had some ability for cue abstraction, in Study 2 we extended our analysis in two ways: First, we compared the predictions of the cue-abstraction model to those of a permutation-based model, in which hypotheses corresponded to a specific order of the four monsters (see Table 2) and, second, we included a sample of 7-year-olds.

Study 1

Method

Participants

Participants were 26 five-year-olds (14 female; $M = 5.59$ years, $SD = .35$ years) and 25 six-year-olds (15 female, $M = 6.45$ years, $SD = .36$ years), recruited and tested at museums and primary schools in the East Bay of the San Francisco area. An additional 17 participants were excluded due to equipment malfunction ($n = 14$), withdrawal of consent ($n = 1$), lack of fluency in English ($n = 1$), or failure to complete the study ($n = 1$). Ethics approval for this study (name: Runners) was obtained by the Institutional Review Board (IRB) of the University of California, Berkeley, and parents gave informed consent for their children's participation before the experiment. The children were native English speakers or fluent in English, were predominantly White and came from various social classes.

Sample size was determined based on standard practices at the time of data collection (2014), which were to include 20 to 30 children per age group, per condition. We note that this method for determining sample sizes is no longer consistent with current best-practices in psychological research, which are that a poweranalysis should be carried out to determine sample size a-priori. However, as such a power analysis requires establishing an effect size of interest based on previous empirical data, and there was a lack of sufficiently similar previous studies and analyses (specifically comparing hypothesis space representations between groups using computational modeling), defining such an effect size would have been extremely speculative.

Hypothesis	Color direction	Shape direction	Cue order
1	B > G	S > C	CO > SH
2	B > G	S > C	SH > CO
3	B > G	C > S	CO > SH
4	B > G	C > S	SH > CO
5	G > B	S > C	CO > SH
6	G > B	S > C	SH > CO
7	G > B	C > S	CO > SH
8	G > B	C > S	SH > CO

Table 1. *Structure of the Cue-Abstraction Model*
Note. B = blue; G = green; S = square; C = circle; CO = color; SH = shape.

Hypothesis	Monster order	Hypothesis	Monster order
1	BC > BS > GC > GS	13	GS > BC > GC > BS
2	BS > BC > GC > GS	14	BC > GS > GC > BS
3	GC > BC > BS > GS	15	GC > GS > BC > BS
4	BC > GC > BS > GS	16	GS > GC > BC > BS
5	BS > GC > BC > GS	17	BC > GC > GS > BS
6	GC > BS > BC > GS	18	GC > BC > GS > BS
7	GC > BS > GS > BC	19	BS > BC > GS > GC
8	BS > GC > GS > BC	20	BC > BS > GS > GC
9	GS > GC > BS > BC	21	GS > BS > BC > GC
10	GC > GS > BS > BC	22	BS > GS > BC > GC
11	BS > GS > GC > BC	23	BC > GS > BS > GC
12	GS > BS > GC > BC	24	GS > BC > BS > GC

Table 2. *Structure of the Permutation-Based Hypothesis Space Model*
Note. GC = green circle; GS = green square; BC = blue circle; BS = blue square.

Furthermore, we decided not to conduct a post hoc power analysis, as it would only be useful with this predetermined effect size and variance. Any other methods essentially represent a reformulation of the stated p-values (O'Keefe, 2007; Thomas, 1997; Zhang et al., 2019) and do not provide any additional information. We believe the fact that we found statistically significant results for most of the hypothesized effects suggests that our studies were sufficiently powered for between-subjects analyses.

Design and Procedure

The task was presented to the children on a laptop computer and consisted of an active learning and a test phase. The task lasted approximately 10 min. The video illustrating the procedure of Study 2, the scripts for both studies and the pseudonymized data can be found at this Open Science Framework (OSF) link: https://osf.io/d75bm/?view_only=ab671227125043d1b7a4d7d0cedddd61 (Jones et al., 2021).

Active Learning Phase

Children were presented with four monsters: a green square, a green round, a blue square, and a blue round monster (see Figure 1). The speed of each monster was determined by its features (i.e., color and shape): The blue monsters were faster than the green ones, the square ones faster than the circular ones, and color was the more important cue for predicting monsters' relative speed. The monster order from the fastest to the slowest was: blue square, blue circle, green square, green circle.

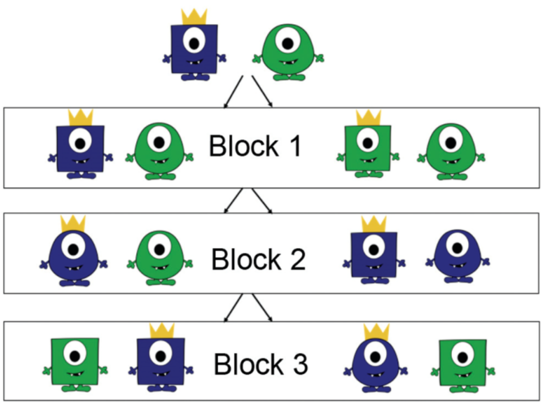

Figure 2. *Pairs of Monsters Presented in the Three Blocks of the Active Learning Phase of Study 1*
Note. The winning monster for each pair is marked with a crown

Participants were first presented with a short video clip showing the blue square monster winning a race against the green circular monster. Verbal instructions prompted participants to pay particular attention to the monsters' color and shape. The active learning phase was carried out in three blocks. In each block, children were presented with two monster pairs on the screen (see Figure 1). Children selected which of the two monster pairs they wanted to see racing to find out which kinds of monsters were faster (Figures 1 and 2). They then saw a video of the selected monsters racing and observed the outcome of each race. Monster pairs that participants had already seen racing appeared on the right side of the screen, with the winner marked with a crown (see Figure 1). These memory aids were provided during the entire learning phase. The learning blocks were designed such that according to the cue-abstraction representation of the hypothesis space, one monster pair provided new information about the cue-outcome relationships, whereas the other provided no new information (EIG = 0; e.g., selecting a blue square monster and a blue circular monster after having seen a green square and green circular monster racing, which only confirms that square monsters are faster, without providing any new information).

If, in any block, a child selected the monster pair that did not provide any information, as outlined above, the same block was presented again. This ensured that children had collected all the relevant information they would need, according to the cue-abstraction representation of the hypothesis space, to make the predictions presented in the test phase. Because of this, children with a higher tendency to select uninformative pairs would have been presented with more learning blocks in total; children who always selected the most informative pair would have been presented with only three learning blocks. However, note that because children in both age groups observed a similar number of training blocks (see the Results section), the systematic differences observed in their performance at test cannot be attributed to the differences in the amount of information they had observed. This design strongly encouraged children to use a cue-abstraction hypothesis space representation through the combination of preselected monster pairs, the repetition of blocks in which children made uninformative choices (according to the cue-abstraction hypothesis space) and the presence of memory aids to reduce the difficulty of the task. With this design, we wanted to test to what extent children can rely on a cue-abstraction representation under the most ideal conditions.

Test Phase

The test phase consisted of a generalization task and a podium task, designed to assess how well children had learned the cue-outcome relationships. In the *generalization* task, children were shown pairs of unfamiliar monsters and were asked to predict which monster would win the race (see Figure 1). Memory aids from the learning phase remained visible on the right side of the screen. The first two trials in the generalization task presented monsters with either known colors but new shapes (blue and green triangles) or known shapes and a new color (a red circle and red square). These two trials aimed to test whether children had learned the direction of the color and shape cues, that is, which color and which shape indicated a higher speed. The order of these trials was counterbalanced across participants. The third trial presented a pair in which one monster had the faster shape but the slower color (a green square) and the other had the slower shape but the faster color (a blue circular monster; see Figure 1). This trial tested whether children had learned the order of the cues, that is, that color was more important than shape for predicting the winner of a race.

In the *podium* task, children were presented with the four monsters from the learning phase and were instructed to rank them from the fastest to the slowest by positioning them on a four-step podium (see Figure 1).

Results

Active Learning Phase

Both age groups showed very similar learning patterns, with comparable mean number of learning blocks ($M_{5years} =$ 5.50, $SD =$ .81; $M_{6years} =$ 5.16, $SD =$ .85; $t(49) =$ 1.46; $p =$ .15) and cue-abstraction-EIG (i.e., EIG calculated according to the cue-abstraction representation of the hypothesis space; $M_{5years} =$ .66, $SD =$ .10; $M_{6years} =$ .71, $SD =$ .12; $t(46.375) = -1.59$, $p =$ .12).

We also found that children's selections were significantly more informative than those of a sample of hypothetical random learners ($t(85.91) = 3.53$, $p <$ .001), although still less informative than those generated by hypothetical optimal learners ($t(50) = -13.34$, $p <$ .001). However, this difference was mostly driven by 6-year-olds ($t(42.17) = 3.37$, $p =$ .002), as the cue-abstraction-EIG of 5-year-olds did not statistically differ from random ($t(40.634) = 1.56$, $p =$ .13). These results suggest that 6-year-olds, but not 5-year-olds, were able to reason following a cue-abstraction representation of the hypothesis space, at least to a certain extent.

Test Phase

If children were able to build up a representation of the hypothesis space based on cue abstraction, they should have known, at the end of the active learning phase, that square monsters were faster than round monsters, that blue monsters were faster than green ones, and that color was more important than shape to determine speed. Note that only information encoded at this level of abstraction was generalizable to new monsters. To quantify children's generalization ability, we calculated the percentage of generalization trials in which children selected the correct monster as predicted by the cue-abstraction model (i.e., the square red monster over the round red monster, the blue monster over the green monster, and the round blue monster over the square green monster). Overall, mean accuracy was significantly above chance for both age groups ($M_{5years} =$ 67.97%, $SD =$ 30.52, exact binomial: $p <$ .001; $M_{6years} =$ 78.67%, $SD =$ 25.24, exact binomial: $p <$ .001).

A logistic regression with continuous age, generalization trial type, mean cue-abstraction-EIG, and the number of learning trials as predictors corroborated the age effect (odds ratio, *OR* = 1.47, [1.19, 1.82], $p <$ .001), and showed that children who did fewer active learning blocks ($OR =$ .76, [.60, .96], $p =$ .023) and whose cue-abstraction-EIG was higher ($OR =$ 5.13, [1.48, 17.75], $p =$ .010) were more likely to correctly predict the winning monsters. The cue order trial was more difficult overall ($OR =$ .42, [.27, .65], $p <$ .001; Nagelkerke's $R^2 =$ .107).

If children had entertained a cue-abstraction representation of the hypothesis space, they would have narrowed down the

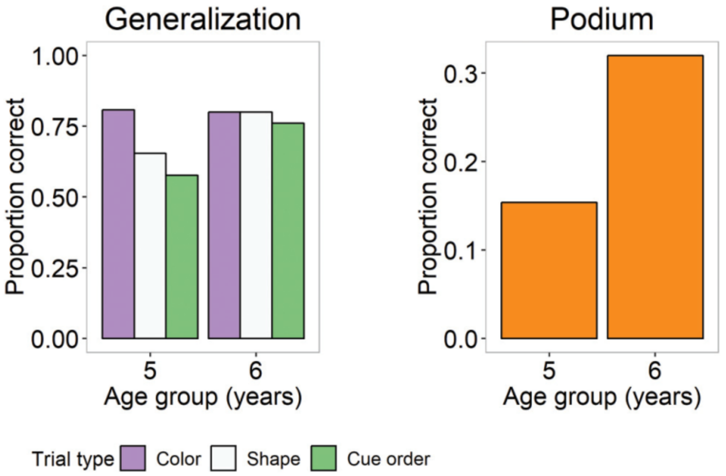

Figure 3. *Performance in the Test Phase of Study 1*
Note. Left: Proportion of children who chose the correct monster on each generalization trial, by age group. Right: Proportion of children who ranked all four monsters correctly on the podium test. With 24 possible rankings, chance performance is equal to. 042.

hypotheses to one by the end of the active learning phase, determining a unique rank order of the monsters. Podium accuracy was scored according to whether children assembled the correct podium, which was compatible with a cue-abstraction representation (1 = correct, 0 = incorrect). Accuracy in the podium task was above chance for both age groups (see Figure 3), with 15.38% of the 5-year-olds (two-tailed binomial test: $p = .02$) and 32% of the 6-year-olds (two-tailed binomial test: $p < .001$) assembling the correct podium. A logistic regression with continuous age as a predictor showed an age effect ($OR = 1.32$, [.01, 2.62], $p = .048$; Nagelkerke's $R^2 = .125$). Adding cue-abstraction-EIG and number of active learning blocks as predictors did not significantly improve model fit ($\chi^2(1) = .098$, $p = .75$, $w = .04$).

Discussion of Study 1

Study 1 provides evidence that, when actively searching, 6-year-olds can, in principle and to some extent, organize their hypothesis space by abstracting the cue-outcome relationships—at least under the conditions of Study 1, which strongly encouraged a cue-abstraction representation. Five-year-olds' selections, by contrast, did not differ from random search. In the generalization and podium tasks, the performance of both age groups differed from chance—suggesting reliance on some systematic approach for solving the task—independent of whether search had been driven by a cue-abstraction representation of the hypothesis space. This suggests that, at test, all children may have organized the information collected during the active learning phase according to a cue-abstraction representation of the hypothesis space. Possibly, even 5-year-olds, who did not seem to rely on a cue-abstraction representation of the hypothesis space when making selections, were nevertheless able to reason about the information collected by abstracting the cue-outcome relationships at test, perhaps because otherwise they would have had nothing to inform their predictions.

Overall, however, children seemed to struggle with abstracting cue order, a second-order cue, as their performance on this generalization trial was lower than on the others. Therefore, while they may have been able to reorganize their hypothesis space at test, young children nonetheless seemed to be limited by the degree of abstraction they can represent. Note that this reorganization was facilitated by the memory aids, presenting all the information observed during the learning phase. Another interpretation of 5-year-olds' seemingly random selections could be that they did organize their hypothesis space by cue-abstraction but were unable to use this organization to direct their search such that each search step maximized EIG. However, our design does not allow us to determine which alternative strategy they could have used.

Since this study encouraged the use of cue-abstraction through the forced-choice format of the active learning phase, its results do not necessarily reflect how children *spontaneously* approach multiple-cue inference tasks. Therefore, we cannot draw any firm conclusions about the strategies and hypothesis space representations they entertain when faced with this type of task. Furthermore, it is also possible that some children tried to use different, irrelevant cues to discern which types of monsters were faster; for example, by observing the relative distance between the monsters as they ran on the track. If this was the case, some children may have struggled to focus on the relevant cues (color and shape) and this may have impacted how well they could reason about these cues and complete the test phase. Study 2 addresses these limitations.

Study 2

To examine how young children spontaneously approach a multiple-cue inference task, in Study 2 we left children free to create their own monster pairs and observe as many races as they wanted. This allowed us to analyze whether children's selections and predictions were better described and compatible with a cue-abstraction (see Table 1) or a permutation-based (see Table 2) representation of the hypothesis space. The two different representations of the hypothesis space make different predictions as to which monster pairs are most informative to observe and should be selected, and also make different predictions for the test phase, because the collected information is encoded differently and narrows down the hypothesis space in different ways. For example, knowing that the green square monster is faster than the green circular monster will leave open four out of eight hypotheses in the cue-abstraction hypothesis space, and 12 out of 24 hypotheses in the permutation-based hypothesis space. In this case, subsequently observing the blue square monster and the blue circular monster racing would not be informative at all in the cue-abstraction model, because it does not rule out any of the four remaining hypotheses. However, the same observation would rule out six out of 12 remaining hypotheses in the permutation-based model, making it the most informative action available.

An additional limitation of Study 1 was that rather than focusing on the relevant cues (color and shape) to assess the relative speed of the monsters, the children could have used other visual cues available during the race, before the monsters crossed the finish line. To address this, in Study 2 we obscured the final half of the race track with trees (see Figure 1) and ensured that each monster ran at the same speed, and was in the same position before reaching the trees, such that a winner could not be inferred from the distance between them. After being hidden by the trees, one monster emerged first to win the race, indicating it must have accelerated while hidden behind the trees. Furthermore, to better capture the developmental trajectory, we extended our age range to include 7-year-olds. Finally, we introduced a memory load manipulation that determined whether children were provided with memory aids (a record of the previous selections and outcomes). The rationale was that this would help us to understand the impact of memory load on children's hypothesis space representation, and to shed some light on the question of which of the two types of representation incurs higher cognitive costs.

Based on the results of Study 1, we expected that older children would be more likely than younger children to rely on a cue-abstraction representation of the hypothesis space during search. We further hypothesized that children would be more likely to represent the hypothesis space in terms of cue-abstraction when assigned to a high memory load condition—where encoding and updating a smaller hypothesis space might be particularly beneficial—than when assigned to a low memory load condition. Finally, based on the results of Study 1, we expected that all children would be more likely to rely on a cue-abstraction representation during the test phase than during the active learning phase, as this type of representation enables more efficient learning and predictions than a permutation-based hypothesis space.

Method

Participants

Participants were 65 five-year-olds (33 female; M = 5.48 years, SD = .32 years), 67 six-year-olds (33 female; M = 6.53 years, SD = .30 years), and 68 seven-year-olds (33 female; M = 7.51 years, SD = .30 years). Sample size was determined based on standard practices at the time of data collection (2017). See the Method section of Study 1 for a discussion of statistical power. A further three participants were excluded because of withdrawal of consent, and two because of technical error. Children were recruited and tested in museums in Berlin, Germany. They were German or fluent in German, were predominantly White and came from various social classes. Ethics approval was obtained by the IRB of the Max Planck Institute for Human Development (study name: Runners 2) and parents gave informed consent for their children to participate before the start of the study.

Design and Procedure

The experiment was presented on a tablet. The paradigm was identical to the one used in Study 1, with the following modifications: First, children were presented with the four monsters and could freely select the learning pairs they wanted to see racing, instead of having to choose between preselected pairs. We incentivized children to search efficiently by initially providing them with 10 stickers, of which they had to "pay" one every time they wanted to see a monster pair racing. They were told that they could keep the stickers left over at the end of the game. Children had to observe at least three monster pairs racing before moving on to the test phase, but could see up to 10 monster races, using all the stickers they were given. After the third trial, they were asked after each race whether they wanted to see another monster pair racing or whether they knew what kinds of monsters were faster and were ready to move on to the test phase. A second modification compared with Study 1 was that the final half of the racetrack was obscured by trees (see Figure 1), to prevent children from basing their predictions on visual cues other than color or shape, such as the distance between the monsters before reaching the finish line. Third, we randomized the correct monster order between participants. Note that all the implemented orderings were potentially consistent with a cue-abstraction representation of the hypothesis space, that is, abstraction of cues was always possible. Fourth, we manipulated memory load by randomly assigning children to one of two conditions: Children in the *high-load* condition did not receive any memory aids, whereas children in the *low-load* condition received memory aids consisting of the previously seen monster pairs being presented on the screen with the winning monster in each pair indicated by a crown (just like in Study 1; see Figure 1). The test phase remained unchanged, except for the third generalization pair, which was slightly altered to make it more distinct from the pairs encountered during the learning phase (see Figure 1).

Results

Active Learning Phase

Aggregated across age groups and conditions, children observed an average of 3.29 (SD = .74) races before moving on to the test phase. The number of learning trials did not differ as a function of continuous age (incidence rate ratio [IRR] = .97, [.89, 1.05], p = .44) or memory load (IRR) = 1.04, [.89, 1.21], p = .64).

	Study 2			
	High load		Low load	
Age group	Cue-abstraction	Permutation	Cue-abstraction	Permutation
5-year-olds	.39 (.15)	.56 (.12)	.45 (.16)	.59 (.10)
6-year-olds	.48 (.12)	.62 (.07)	.48 (.13)	.61 (.10)
7-year-olds	.53 (.12)	.62 (.09)	.48 (.11)	.64 (.07)

Table 3. Mean Expected Information Gain (and SD) of Selections Calculated According to the Cue-Abstraction and Permutation-Based Representations of the Hypothesis Space, Displayed by Age Group and Condition
Note. GC = green circle; GS = green square; BC = blue circle; BS = blue square.

We calculated the EIG of each child's selections according to both the cue-abstraction and the permutation-based representations of the hypothesis space. Overall, the average EIG of children's selections was lower when calculated according to the cue-abstraction than when calculated according to the permutation-based representation of the hypothesis space (see Table 3). In itself, however, this does not provide conclusive evidence that the children were likely to have relied on a permutation-based representation. Due to the fact that the hypothesis space with a permutation-based representation is larger than with a cue-abstraction representation, any query is usually informative to some extent for the former, leading to a higher overall EIG under the permutation-based than under the cue-abstraction model, where only a few queries are informative at each step. The latent mixture-model analysis presented below provides a more diagnostic comparison of the two models. To foreshadow the results, these analyses suggest—consistent with the EIG analysis—that children's search was better captured by a permutation-based representation of the hypothesis space.

A linear regression with continuous age, memory load, and their interaction as predictors showed that cue-abstraction-EIG increased with age ($\beta = .10$, [.06, .14], $p < .001$) and was higher in the low memory load than in the high memory load condition ($\beta = .59$, [.20, .97], $p = .003$). There was also a significant Age × Memory Load interaction ($\beta = -.09$, [$-.15$, $-.03$], $p = .004$; $F(3, 196) = 7.597$, $p < .001$; $R^2 = .09$; $f^2 = .12$), indicating that the effect of memory load on mean cue-abstraction-EIG decreased with age. A linear regression of permutation-based-EIG also showed a main effect of age ($\beta = .04$, [.02, .06], $p < .001$) but not of memory load ($\beta = .02$, [$-.02$, .05], $p = .32$; $F(2, 197) = 6.888$; $p = .001$; $R^2 = .056$; $f^2 = .07$). Adding an Age × Memory Load interaction did not significantly improve model fit ($\chi^2(1) = .016$, $p = .32$, $w = .009$).

Children often stopped the search before they had narrowed down the hypothesis space to a single hypothesis, that is, when there was still information to be gained according to either representation of the hypothesis space. Although the prevalence of such early stopping declined with age (5-year-olds: 67.69%; 6-year-olds: 67.16%; 7-year-olds: 54.41%), a logistic regression showed that these differences were not significant ($OR = .74$, [.53, 1.04], $p = .081$) and that rates of early stopping were unaffected by memory load condition ($OR = 1.01$, [.56, 1.80], $p = .981$, Nagelkerke's $R^2 = .021$). This indicates that children did not markedly differ in their ability to search exhaustively.

Test Phase

As mentioned in the results section of Study 1, children's predictions in the generalization trials would be much more accurate if they had encoded the information collected in the active learning phase according to a cue-abstraction, compared with a permutation-based representation of the hypothesis space. Aggregating across conditions, performance in the generalization task was above chance only for 6- and 7-year-olds ($M_{5years} = 57\%$, $SD = 32$, two-tailed binomial test: $p = .06$; $M_{6years} = 69\%$, $SD = 28$, $p < .001$; $M_{7years} = 76\%$, $SD = 25$, $p < .001$). A logistic regression showed that the probability of making a correct prediction increased significantly with age (Figure 4; $OR = 1.47$, [1.19, 1.82], $p < .001$), but there was no significant effect of memory load ($OR = 1.28$, [.89, 1.84], $p = .175$). As observed in Study 1, the probability of responding correctly in the cue order trial was overall lower than on the other trials (Figure 4; $OR = .42$, [.27, .65], $p < .001$). A higher number of learning trials was negatively associated with the probability of giving a correct response ($OR = .76$, [.60, .96], $p = .023$).

Like in Study 1, we found that a higher cue-abstraction-EIG during search increased the probability of giving a correct response ($OR = 5.13$, [1.48, 17.75], $p = .01$) in the generalization trials. As expected, the EIG achieved under the permutation-based hypothesis space was unrelated to on the probability of giving a correct response ($OR = .20$, [.03, 1.24], $p = .084$; Nagelkerke's $R^2 = .107$). This suggests that those children who were already searching guided by a cue-abstraction representation of the hypothesis space

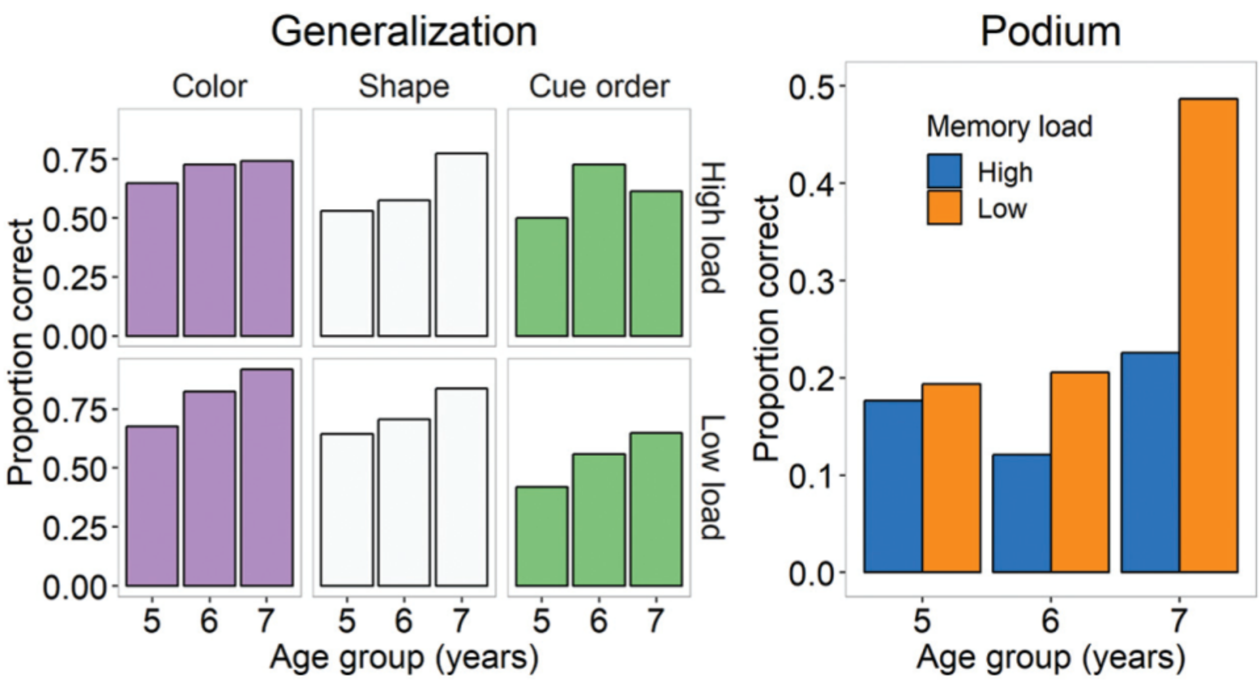

Figure 4. *Performance in the Test Phase of Study 2*
Note. Left: Proportion of participants who responded correctly in the generalization task for each trial type (vertical panels) in high and low memory load (horizontal panels). Right: Proportion of participants who got the podium composition correct in each condition.

were also more likely to rely on this representation during the generalization trials.

In the podium task, the proportion of children who indicated the correct composition of the podium increased with age and was above chance level at all ages (see Figure 4): 18.46% of 5-year-olds, 16.42% of 6-year-olds, and 36.76% of 7-year-olds identified the correct podium. We used logistic regression to assess to what extent age and memory load were related to performance in the podium task. The probability of assembling a correct podium increased significantly with age ($OR = 1.71$, 95% confidence interval, CI [1.15, 2.54], $p = .008$) and in the low-load condition ($OR = 2.02$, 95% CI [1.02, 4.00], $p = .044$, Nagelkerke's $R^2 = .089$). We also tested a regression model that included an Age × Condition interaction, but this did not significantly improve model fit ($\chi^2(1) = 1.41$, $p = .24$, $w = .08$). Adding EIG achieved during learning according to both hypothesis space models and the number of learning trials as predictors also failed to improve model fit ($\chi^2(3) = 2.073$, $p = .55$, $w = .10$).

Mixture-Model Comparison Across Active Learning and Test Phases

The results above suggest that children may have relied on different representations of the hypothesis space during search (i.e., learning) than during prediction, in that selections during the active learning phase seemed to be more consistent with a permutation-based hypothesis space, whereas performance during the test showed evidence of cue abstraction, particularly among older children.

To directly compare reliance on cue-abstraction versus permutation-based hypothesis spaces, we modeled children's selections and predictions using a hierarchical Bayesian latent mixture model (Bartlema et al., 2014; Bramley et al., 2015; see online supplementary materials for full details. Supplementary material to this article is available. For more information see http://hdl.handle.net/21.11116/0000-0007-1E82-5). The mixture model compared three candidate models by estimating their probability of having generated participants' choices in each phase: a random model (RAND), that corresponded to random selection during the learning phase and guessing in the test phase; the permutation-based model (PERM) and the cue-abstraction model (CA). Both the PERM and CA hypothesis spaces assume that selections during the active learning phase are driven by an EIG-maximizing strategy (as calculated using the corresponding hypothesis space; see Supplementary materials S1) and the resulting outcomes are used to rule out hypotheses that are inconsistent with the evidence. The set of hypotheses that remain at the end of the search phase are then used to predict responses on the generalization and podium tests. The estimated mixture probabilities (θ) indicate the likelihood of each model within each group based on how well that model fits participants' choices.

This computational modeling approach complements the behavioral analyses in a number of ways. First, it provides an integrative account of hypothesis space use during the test phase, as all of an individual's responses (generalization trials and podium) are modeled as arising from a common underlying representation of the hypothesis space. Second, it allows us to directly weigh the evidence for each representation when taking into account the information collected by each individual during learning. This is important because participants selected different sets of observations during the active learning phase, which in turn affects the expected performance on test trials. For example, the PERM and CA hypothesis space models both predict low test accuracy among participants who make uninformative selections and end the learning phase without narrowing down either hypothesis space. Similarly, a small number of high-EIG observations may be enough to identify the correct hypothesis under the CA model, but still leave uncertainty under the PERM model. Examining the estimated mixture probabilities provides a clearer indication of a group's reliance on each hypothesis space representation, given all observed behaviors in the active learning and the test phase.

Figure 5 shows the posterior means and 95% highest posterior density intervals for the mixture probabilities θ in the learning phase (top row) and the test phase (bottom row). Confirming the EIG results presented above, all children's selections in the learning phase were better captured by the PERM model (see Supplemental Table S1 for pairwise differences between mixture probabilities), with one exception: For 5-year-olds in the high-load condition, the RAND model had the highest posterior probability. This suggests that the task demands may have been too high for the youngest age group to implement a systematic search strategy or hypothesis space representation. In the test phase, which jointly considers the responses to the generalization and the podium tests, the hypothesis space representation seemed to be less consistent for all age groups (see Supplemental Table S1 for pairwise comparisons between strategies). In contrast to the behavioral results,

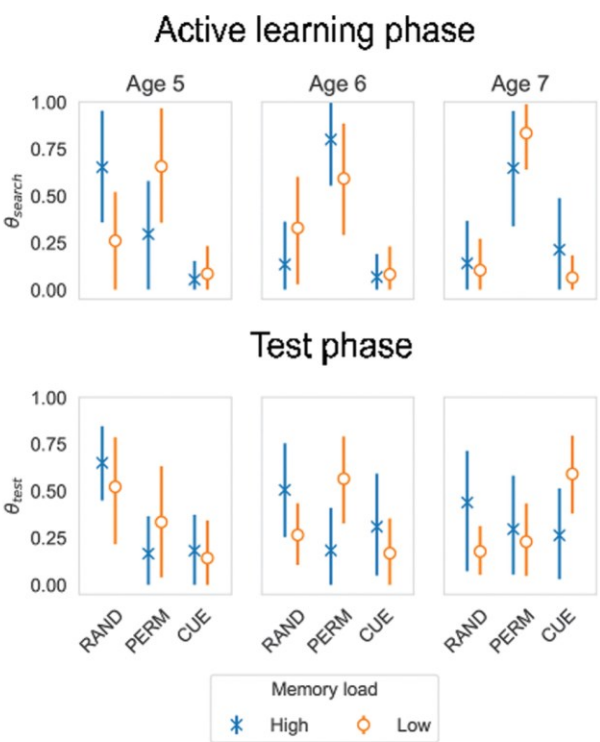

Figure 5. *Mixture Probabilities of Each Strategy in Study 2 for the Learning Phase (Top) and Test Phase (Bottom)*
Note. Error bars represent 95% highest posterior density intervals.

the RAND model best captured 5-year-olds' predictions in both memory load conditions, as well as 6- and 7-year-olds' predictions in the high-load condition. In the low-load condition, the PERM model was more likely for the 6-year-olds, whereas the CA model was more likely among the 7-year-olds.

Discussion of Study 2

In Study 2, we compared two possible representations of the hypothesis space—cue-abstraction and permutation-based—in their ability to capture children's selections and predictions in an active learning multiple-cue inference task. We also considered a model that assumed no systematic approach, and that would result in a random selection during search and guessing at test.

The results of Study 2 build upon and extend the findings of Study 1. We confirmed that, although there are indications that children are to some extent able to abstract cue-outcome relationships, they struggle to abstract cue order, as shown by their lower performance in the cue order generalization trial. Instead, the 6- and 7-year-olds represented the hypothesis space in terms of a permutation-based representation. This held for both the active learning as well as—for the 6-year-olds—the test phase in the low memory load condition. The 7-year-olds, however, seemed to *shift* to a representation based on cue abstraction in the test phase, suggesting that this representation is more demanding. Consistent with this view, and contrary to our hypothesis, there was no longer any evidence for reliance on a cue-abstraction representation in the high memory load condition. Instead, 6- and 7-year-olds were better fit by a random model. This suggests that, even when their search was guided by a permutation-based representation of the hypothesis space, integrating the information collected and making informed predictions without memory aids was too challenging. Five-year-olds were generally better captured by a random model, indicating that the task might have been too challenging for them, even when provided with memory aids.

Assuming that children's hypothesis space representations were constrained by working memory capacity, these results suggest that the permutation-based representation may incur lower costs. This is consistent with Hoffmann et al. (2013), as well as the indications that working memory and executive functions undergo significant developmental changes between the ages of 4 and 14 (reviewed in Best & Miller, 2010). Accordingly, 5-year-olds' executive functions may not have been sufficiently developed to support a systematic organization of the hypothesis space, but in 6-year-olds they are developed enough to support the coordination of working memory subcomponents (Best & Miller, 2010), hence their use of a permutation-based strategy. Seven-year-olds' executive functions appear to be able to support the more complex cue-abstraction hypothesis space, as long as they are provided with some scaffolding.

As explained above, the results of Studies 1 and 2 suggest that under high memory load—that is, in the absence of memory aids—young children struggle to abstract cue-outcome relationships, and in the case of 5-year-olds, to entertain any systematic representation of the hypothesis space at all. Therefore, children between the ages of 5 and 7 do not yet seem to have sufficiently mature executive functions to spontaneously maintain a systematic hypothesis space representation throughout an entire task without scaffolding. This also points to cognitive factors such as working memory capacity being crucial moderators of how children can represent task-relevant information, possibly driving the observed developmental shifts in information search strategies. However, the results of the active learning phase across the two studies cannot be directly compared, as the forced-choice design in Study 1 strongly encouraged a cue-abstraction representation of the hypothesis space and did not reflect how young children spontaneously approach a multiple-cue inference task.

Additionally, in Study 1 children were forced to continue training until they had seen all of the informative races (according to the cue-abstraction representation of the hypothesis space), so that they had all the information needed for abstracting the cue-outcome relationships. Since many children in Study 2 did not search efficiently, it is also possible that children did not collect enough information for cue abstraction, which may have contributed to their difficulty in maintaining a consistent hypothesis space representation throughout the whole task. However, because we were able to successfully differentiate between the two key hypothesis space models, and because the number of learning trials did not differ between age groups—excluding the possibility that one group systematically failed to collect enough information compared with the other—it seems unlikely that the children had collected too little information to construct a coherent hypothesis space. This difficulty in *generating* versus *selecting* the most informative actions is in line with developmental findings from question-asking research (Ruggeri et al., 2016, 2017; Ruggeri & Feufel, 2015). This work shows that children younger than 10 years of age struggle to spontaneously generate the most informative questions, although they can already select the most informative questions at age 5.

Although the memory load manipulation did not affect children's search patterns or generalization performance, their accuracy in the podium task was higher under low than under high memory load (particularly for the 7-year-olds). This indicates that although higher memory load made it more difficult to learn the monster orders seen during learning, it did not prevent (at least) older children from making generalizations to new objects. All children performed above chance level in the podium task, suggesting that the memory load did not make learning impossible (even for 5-year-olds). Note that this conclusion seems to be at odds with the mixture-modeling results, according to which 5-year-olds were best captured by a random approach to both information search and test predictions, and 6-year-olds were only able to use an organized hypothesis space (the permutation-based representation) under low load. The modeling results would suggest chance-level performance for 5-year-olds in both conditions and 6-year-olds in the high-load condition; in addition, for the 6-year-olds generalization accuracy should have been lower in the low-load condition. The fact that, behaviorally, their performance was reliably above chance level indicates that these children were either able to memorize the monster orders they saw during the learning phase, or that they used a strategy that we did not consider in our analysis. We also note that while children's behavioral results were consistent with the intended effect of our memory load manipulation, we could not precisely measure exactly how much load differed between conditions. Future research should address this limitation, confirming and exploring the effects of memory load on hypothesis space representations in more detail.

General Discussion

We presented two studies in which we explored how 5- to 7-year-olds actively search for information to make accurate predictions based on multiple cues. Moreover, using computational modeling, we investigated what children's search patterns and predictions revealed about their representation of the hypothesis space. The results of Study 1 suggest that 5-year-olds are already able to abstract cue-outcome relationships and generalize to new exemplars, at least as long as scaffolding is given during the active learning phase. This is a much younger age than typically assumed. Indeed, previous work suggested that children have difficulty identifying relevant cues (Betsch et al., 2016; Davidson, 1991b; Montanelli, 1972) and reasoning about hypotheses in an abstract, hierarchical manner until late childhood (i.e., around age 10; Ruggeri & Feufel, 2015). Furthermore, even 9- to 11-year-olds seemed to have difficulty representing cue relationships hierarchically (Von Helversen et al., 2010). To see that 5-year-olds are able to do so, even if only with extensive scaffolding, is a surprising result.

However, Study 2 showed that even the older children in our sample needed scaffolding to support a cue-abstraction hypothesis space, echoing results from studies of question asking (Ruggeri et al., 2017; Ruggeri & Feufel, 2015). Crucially, both studies provided the first evidence that children sometimes *switch* between different representations of the hypothesis space: In some cases, any organization of the hypothesis space that seems to have driven search seems to dissolve at test, with children performing at random. In other cases, children seem to be have been able to spontaneously reorganize the hypothesis space after the search phase (if memory aids are provided), to encode the information collected by abstracting the cue-outcome relationship, supporting their ability to make more accurate predictions. For example, although during the learning phase the 7-year-olds in Study 2 seemed to rely on a permutation-based representation of the hypothesis space (when memory aids were provided), they shifted to a cue-abstraction representation during the test phase (for a related switch in reliance on cue abstraction and exemplar processing between different judgment tasks in adults, see Pachur & Olsson, 2012; Trippas & Pachur, 2019).

More generally, considering how children may represent the hypothesis space at different ages and at different stages of the learning process could help explain inconsistencies in existing findings. For instance, it is well established that the efficiency of children's questions increases dramatically between the ages of 5 and 10, with particularly large improvements from age seven (e.g., Herwig, 1982; Mosher & Hornsby, 1966; Ruggeri et al., 2016). However, the factors driving these changes are still poorly understood. While it seems obvious that the maturation of children's verbal skills and executive functions should be important contributors, another intriguing possibility is that the development of these cognitive skills is tied to or constrains the emergence of the ability to represent relevant information in a hierarchical manner. Being able to entertain such hierarchical representations, together with the flexibility to reorganize the hypothesis space during the course of a task, may be crucial for more effectively guiding information search and supporting prediction. However, previous attempts to relate cognitive skills such as verbal knowledge and short-term memory (STM) to strategy choices and performance in multiple-cue inference tasks have not revealed any strong relationships (Mata et al., 2011). The relationship between developmental shifts in information search strategies, hypothesis space structures, and cognitive skills appears to be complex and calls for more research.

Similarly, achieving a better understanding of how children's search strategies relate to their representation of the hypothesis space could help explain why interventions aimed at improving children's learning strategies do not always work as intended. For example, attempts to improve the efficiency of children's questions (e.g., Courage, 1989; D. R. Denney, 1972; N. W. Denney & Turner, 1979) and use of unconfounded experiments in causal learning (e.g., Chase & Klahr, 2017; Dean & Kuhn, 2007) have met with limited success. If students have misunderstandings about the relevant variables or cues, which should be reflected in their hypothesis space structure, interventions which do not address these misconceptions are unlikely to be effective. Furthermore, if children use search strategies that cannot be directly mapped onto their hypothesis space representations, it would be important to identify these strategies and investigate how and to what extent the hypothesis space guides children's search decisions. Indeed, although our design did not allow us to properly distinguish between different search strategies and hypothesis space representations other than cue-abstraction and permutation-based, because it required us to assume an EIG-maximizing search strategy to compare the candidate hypothesis space models, the results of Study 2 also suggest that some children may have used a search strategy that did not aim to maximize EIG. For instance, other search strategies that have been considered in the multiple-cue inference literature are a Weighted-Additive Rule (WADD; Mata et al., 2011; Parpart et al., 2015; Payne et al., 1993), which integrates all cue values weighted by their importance, a Tally strategy, which assigns equal weights to cues and adds them up (Gigerenzer & Goldstein, 1996; Mata et al., 2011) or a Take-The-Best strategy, which bases a decision on a single cue (Fechner et al., 2018; Gigerenzer & Goldstein, 1996; Mata et al., 2011; Parpart et al., 2015).

Future research might also consider subsets or approximations of the hypothesis space representations we focused on in this article, such as a simpler variant of the cue-abstraction hypothesis space in which only the cue directions are encoded, or a permutation-based hypothesis space with a single hypothesis that children update with each observation. Sequential hypothesis testing with even a single hypothesis can require fewer cognitive resources while being consistent with the goal of efficiently reducing uncertainty (Bonawitz et al., 2014). Finally, a more general approach to studying children's information search strategies could be to use the approach taken by Schulz et al. (2019) to study children and adults' exploration and generalization behavior in a multiarmed bandit task, which involved a computational model with free parameters that allows one to capture participants' behavior with minimal assumptions.

Considering the wider context of the development of the skills needed for multiple-cue inference, our results are also consistent with evidence that children's ability to integrate information efficiently improves with age, though we tested a younger age range than the typical 9- to 13-year-olds considered in previous work (e.g., Kuhn et al., 2008; Mata et al., 2011; Von Helversen et al., 2010). Our findings also provide some support to previous results showing that younger children (9- to 10-year-olds) may prefer information-intensive search strategies that do not require selective attention or complex cue integration (Bereby-Meyer et al., 2004; Davidson, 1991a, 1991b, 1996; Mata et al., 2011), which bear some similarity to the permutation-based hypothesis space representation we considered. Furthermore, the use of such information-intensive strategies is tied to cognitive skills, such that lower cognitive skills favor the use of these kinds of strategies over more information-frugal ones (Mata et al., 2007). Specifically, information-frugal strategies often involve controlled search processes that are cognitively rather effortful (Fechner et al., 2018) and also depend on knowledge about the cue hierarchy, which more naïve decision makers might not have available (Pachur & Marinello, 2013). This is also consistent with our findings and the idea that using cue-abstraction incurs higher cognitive costs than a permutation-based representation.

However, the fact that 5- to 7-year-olds in our studies appeared to be able to use a cue-abstraction hypothesis space with scaffolding suggests that the ability to effectively integrate cue information may emerge earlier than previously thought. We note that previous work has mostly used paradigms in which some cues were more relevant than others, which was not the case in our studies. Therefore, there was no need to selectively attend to some cues and ignore others, nor was there a need to represent cues based on their validity, which may have made the task easier and revealed young children's previously unknown competence. Earlier findings showing 9- and 10-year-olds' struggling to integrate and selectively attend to cue information may have reflected the increased difficulty of the tasks they were given and underestimated their abilities.

In conclusion, by using an active learning paradigm to assess young children's information search and representation of the hypothesis space, we demonstrated that when provided with scaffolding, children as young as 7 years of age show some ability to spontaneously represent a hypothesis space in terms of abstracted cue-outcome relationships—at an earlier age than expected based on previous work on question asking and categorization. Our use of an active learning approach allowed us to investigate the structure of children's hypothesis space more directly and provided the first evidence that participants may not necessarily rely on the same hypothesis space throughout a task. Our findings have potential implications for investigations relating the development of children's information search to their cognitive skills, as well as for some educational interventions. For instance, it may be possible in future research to infer from children's questions and search selections possible misconceptions they might have and help tailor educational interventions to individual students. Future work should investigate whether young children use other types of hypothesis space representations than those we considered, the extent to which this representation changes across tasks and guides information search, and how specific cognitive abilities relate to the use of specific hypothesis space structures.

References

Bartlema, A., Lee, M., Wetzels, R., & Vanpaemel, W. (2014). A Bayesian hierarchical mixture approach to individual differences: Case studies in selective attention and representation in category learning. *Journal of Mathematical Psychology*, *59*, 132–150. https://doi.org/10.1016/j.jmp.2013.12.002

Bereby-Meyer, Y., Assor, A., & Katz, I. (2004). Children's choice strategies: The effects of age and task demands. *Cognitive Development*, *19*(1), 127–146. https://doi.org/10.1016/j.cogdev.2003.11.003

Best, J. R., & Miller, P. H. M. (2010). A developmental perspective on executive function. *Child Development*, *81*(6), 1641–1660. https://doi.org/10.1111/j.1467-8624.2010.01499.x

Betsch, T., Lang, A., Lehmann, A., & Axmann, J. M. (2014). Utilizing probabilities as decision weights in closed and open information boards: A comparison of children and adults. *Acta Psychologica*, *153*, 74–86. https://doi.org/10.1016/j.actpsy.2014.09.008

Betsch, T., Lehmann, A., Lindow, S., Lang, A., & Schoemann, M. (2016). Lost in search: (Mal-)adaptation to probabilistic decision environments in children and adults. *Developmental Psychology*, *52*(2), 311–325. https://doi.org/10.1037/dev0000077

Bonawitz, E. B., Denison, S., Gopnik, A., & Griffiths, T. L. (2014). Win-stay, lose-sample: A simple sequential algorithm for approximating Bayesian inference. *Cognitive Psychology*, *74*, 35–65. https://doi.org/10.1016/j.cogpsych.2014.06.003

Bonawitz, E. B., van Schijndel, T. J. P., Friel, D., & Schulz, L. (2012). Children balance theories and evidence in exploration, explanation, and learning. *Cognitive Psychology*, *64*(4), 215–234. https://doi.org/10.1016/j.cogpsych.2011.12.002

Bramley, N. R., Lagnado, D. A., & Speekenbrink, M. (2015). Conservative forgetful scholars: How people learn causal structure through sequences of interventions. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *41*(3), 708–731. https://doi.org/10.1037/xlm0000061

Chase, C. C., & Klahr, D. (2017). Invention versus direct instruction: For some content, it's a tie. *Journal of Science Education and Technology*, *26*(6), 582–596. https://doi.org/10.1007/s10956-017-9700-6

Cook, C., Goodman, N. D., & Schulz, L. E. (2011). Where science starts: Spontaneous experiments in preschoolers' exploratory play. *Cognition*, *120*(3), 341–349. https://doi.org/10.1016/J.COGNITION.2011.03.003

Courage, M. L. (1989). Children's inquiry strategies in referential communication and in the game of twenty questions. *Child Development*, *60*(4), 877–886. https://doi.org/10.2307/1131029

Davidson, D. (1991a). Children's decision-making examined with an information-board procedure. *Cognitive Development*, *6*(1), 77–90. https://doi.org/10.1016/0885-2014(91)90007-Z

Davidson, D. (1991b). Developmental differences in children's search of predecisional information. *Journal of Experimental Child Psychology*, *52*(2), 239–255. https://doi.org/10.1016/0022-0965(91)90061-V

Davidson, D. (1996). The effects of decision characteristics on children's selective search of predecisional information. *Acta Psychologica*, *92*(3), 263–281. https://doi.org/10.1016/0001-6918(95)00014-3

Dean, D. J., & Kuhn, D. (2007). Direct instruction vs. discovery: The long view. *Science Education*, *91*(3), 384–397. https://doi.org/10.1002/sce.20194

Denney, D. R. (1972). Modeling and eliciting effects upon conceptual strategies. *Child Development*, *43*(3), 810–823. https://doi.org/10.2307/1127633

Denney, N. W., & Turner, M. C. (1979). Facilitating cognitive performance in children: A comparison of strategy modeling and strategy modeling with overt self-verbalization. *Journal of Experimental Child Psychology*, *28*(1), 119–131. https://doi.org/10.1016/0022-0965(79)90106-1

Einhorn, H. J., Kleinmuntz, D. N., & Kleinmuntz, B. (1979). Linear regression and process-tracing models of judgment. *Psychological Review*, *86*(5), 465–485. https://doi.org/10.1037/0033-295X.86.5.465

Enkvist, T., Newell, B., Juslin, P., & Olsson, H. (2006). On the role of causal intervention in multiple-cue judgment: Positive and negative effects on learning. *Journal of Experimental Psychology: Learning Memory & Cognition*, *32*(1), 163–179. https://doi.org/10.1037/0278-7393.32.1.163

Fechner, H. B., Schooler, L. J., & Pachur, T. (2018). Cognitive costs of decision-making strategies: A resource demand decomposition analysis

with a cognitive architecture. *Cognition*, *170*, 102–122. https://doi.org/10.1016/j.cognition.2017.09.003
Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review*, *103*(4), 650–669. https://doi.org/10.1093/acprof
Gregan-Paxton, J., & John, D. R. (1995). Are young children adaptive decision makers? A study of age differences in information search behavior. *Journal of Consumer Research*, *21*(4), 567–580. https://doi.org/10.1086/209419
Gregan-Paxton, J., & John, D. R. (1997). The emergence of adaptive decision making in children. *Journal of Consumer Research*, *24*(1), 43–56. https://doi.org/10.1086/209492
Herwig, J. E. (1982). Effects of age, stimuli, and category recognition factors in children's inquiry behavior. *Journal of Experimental Child Psychology*, *33*(2), 196–206. https://doi.org/10.1016/0022-0965(82)90015-7
Hoffmann, J. A., von Helversen, B., & Rieskamp, J. (2013). Deliberation's blindsight: How cognitive load can improve judgements. *Psychological Science*, *24*(6), 869–879. https://doi.org/10.1177/0956797612463581
Horn, S. S., Ruggeri, A., & Pachur, T. (2016). The development of adaptive decision making: Recognition-based inference in children and adolescents. *Developmental Psychology*, *52*(9), 1470–1485. https://doi.org/10.1037/dev0000181
Howse, R. B., Best, D. L., & Stone, E. R. (2003). Children's decision making: The effects of training, reinforcement, and memory aids. *Cognitive Development*, *18*(2), 247–268. https://doi.org/10.1016/S0885-2014(03)00023-6
Jones, A., Markant, D. B., Pachur, T., Gopnik, A., & Ruggeri, A. (2021). *How is the hypothesis space represented? Evidence from young children's active search and predictions in a multiple-cue inference task*. https://doi.org/10.17605/OSF.IO/D75BM
Juslin, P., Jones, S., Olsson, H., & Winman, A. (2003). Cue abstraction and exemplar memory in categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(5), 924–941. https://doi.org/10.1037/0278-7393.29.5.924
Juslin, P., Olsson, H., & Olsson, A.-C. (2003). Exemplar effects in categorization and multiple-cue judgment. *Journal of Experimental Psychology: General*, *132*(1), 133–156. https://doi.org/10.1037/0096-3445.132.1.133
Kuhn, D., Iordanou, K., Pease, M., & Wirkala, C. (2008). Beyond control of variables: What needs to develop to achieve skilled scientific thinking? *Cognitive Development*, *23*(4), 435–451. https://doi.org/10.1016/j.cogdev.2008.09.006
Legare, C. H., Mills, C. M., Souza, A. L., Plummer, L. E., & Yasskin, R. (2013). The use of questions as problem-solving strategies during early childhood. *Journal of Experimental Child Psychology*, *114*(1), 63–76. https://doi.org/10.1016/j.jecp.2012.07.002
Markant, D. B., & Gureckis, T. M. (2014). Is it better to select or to receive? Learning via active and passive hypothesis testing. *Journal of Experimental Psychology: General*, *143*(1), 94–122. https://doi.org/10.1037/a0032108
Markant, D. B., Settles, B., & Gureckis, T. M. (2016). Self-directed learning favors local, rather than global, uncertainty. *Cognitive Science*, *40*(1), 100–120. https://doi.org/10.1111/cogs.12220
Mata, R., Schooler, L. J., & Rieskamp, J. (2007). The aging decision maker: Cognitive aging and the adaptive selection of decision strategies. *Psychology and Aging*, *22*(4), 796–810. https://doi.org/10.1037/0882-7974.22.4.796
Mata, R., von Helversen, B., & Rieskamp, J. (2011). When easy comes hard: The development of adaptive strategy selection. *Child Development*, *82*(2), 687–700. https://doi.org/10.1111/j.1467-8624.2010.01535.x
McCormack, T., Bramley, N., Frosch, C., Patrick, F., & Lagnado, D. A. (2016). Children's use of interventions to learn causal structure. *Journal of Experimental Child Psychology*, *141*, 1–22. https://doi.org/10.1016/j.jecp.2015.06.017
Meder, B., Nelson, J. D., Jones, M., & Ruggeri, A. (2019). Stepwise versus globally optimal search in children and adults. *Cognition*, *191*, Article 103965. https://doi.org/10.1016/j.cognition.2019.05.002
Montanelli, D. S. (1972). Multiple-cue learning in children. *Developmental Psychology*, *7*(3), 302–312. https://doi.org/10.1037/h0033346
Mosher, F. A., & Hornsby, J. R. (1966). On asking questions. In J. S. Bruner, R. R. Oliver, & P. M. Greenfield (Eds.), *Studies in cognitive growth* (pp. 86–102). Wiley.
O'Keefe, D. J. (2007). Brief report: Post hoc power, observed power, a priori power, retrospective power, prospective power, achieved power: Sorting out appropriate uses of statistical power analyses. *Communication Methods and Measures*, *1*(4), 291–299. https://doi.org/10.1080/19312450701641375
Pachur, T., & Marinello, G. (2013). Expert intuitions: How to model the decision strategies of airport customs officers?*Acta Psychologica*, *144*(1), 97–103. https://doi.org/10.1016/j.actpsy.2013.05.003
Pachur, T., & Olsson, H. (2012). Type of learning task impacts performance and strategy selection in decision making. *Cognitive Psychology*, *65*(2), 207–240. https://doi.org/10.1016/j.cogpsych.2012.03.003
Parpart, P., Schulz, E., Speekenbrink, M., & Love, B. C. (2015). Active learning as a means to distinguish among prominent decision strategies. In D. C. Noelle (Eds.), *Proceedings of the 37th Annual Meeting of the Cognitive Science Society* (pp. 1–6). Cognitive Science Society.
Payne, J. W., Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). *The adaptive decision maker*. Cambridge University Press. https://doi.org/10.1017/CBO9781139173933
Ruggeri, A., & Feufel, M. A. (2015). How basic-level objects facilitate question-asking in a categorization task. *Frontiers in Psychology*, *6*, Article 918. https://doi.org/10.3389/fpsyg.2015.00918
Ruggeri, A., & Katsikopoulos, K. V. (2013). Make your own kinds of cues: When children make more accurate inferences than adults. *Journal of Experimental Child Psychology*, *115*(3), 517–535. https://doi.org/10.1016/j.jecp.2012.11.007
Ruggeri, A., & Lombrozo, T. (2015). Children adapt their questions to achieve efficient search. *Cognition*, *143*, 203–216. https://doi.org/10.1016/j.cognition.2015.07.004
Ruggeri, A., Lombrozo, T., Griffiths, T. L., & Xu, F. (2016). Sources of developmental change in the efficiency of information search. *Developmental Psychology*, *52*(12), 2159–2173. https://doi.org/10.1037/dev0000240
Ruggeri, A., Sim, Z. L., & Xu, F. (2017). Why is Toma late to school again?" Preschoolers identify the most informative questions. *Developmental Psychology*, *53*(9), 1620–1632. https://doi.org/10.1037/dev0000340
Ruggeri, A., Swaboda, N., Sim, Z. L., & Gopnik, A. (2019). Shake it baby, but only when needed: Preschoolers adapt their exploratory strategies to the information structure of the task. *Cognition*, *193*, Article 104013. https://doi.org/10.1016/j.cognition.2019.104013
Ruggeri, A., Walker, C., Lombrozo, T., & Gopnik, A. (2021). How to help young children ask better questions. *Frontiers in Psychology*, *11*, Article 2908. https://doi.org/10.3389/fpsyg.2020.586819
Schulz, E., Wu, C. M., Ruggeri, A., & Meder, B. (2019). Searching for rewards like a child means less generalization and more directed exploration. *Psychological Science*, *30*(11), 1561–1572. https://doi.org/10.1177/0956797619863663
Schulz, L. E., & Bonawitz, E. B. (2007). Serious fun: Preschoolers engage in more exploratory play when evidence is confounded. *Developmental Psychology*, *43*(4), 1045–1050. https://doi.org/10.1037/0012-1649.43.4.1045
Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, *27*(3), 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x
Sim, Z., & Xu, F. (2017). Infants preferentially approach and explore the unexpected. *British Journal of Developmental Psychology*, *35*(4), 596–608. https://doi.org/10.1111/bjdp.12198

Smith, E. E., Patalano, A. L., & Jonides, J. (1998). Alternative strategies of categorization. *Cognition*, *65*(2–3), 167–196. https://doi.org/10.1016/S0010-0277(97)00043-7

Stahl, A. E., & Feigenson, L. (2015). Observing the unexpected enhances infants' learning and exploration. *Science*, *348*(6230), 91–94. https://doi.org/10.1126/science.aaa3799

Thomas, L. (1997). Retrospective power analysis. *Conservation Biology*, *11*(1), 276–280. https://doi.org/10.1046/j.1523-1739.1997.96102.x

Trippas, D., & Pachur, T. (2019). Nothing compares: Unraveling learning task effects in judgment and categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*(12), 2239–2266. https://doi.org/10.1037/xlm0000696

Von Helversen, B., Mata, R., & Olsson, H. (2010). Do children profit from looking beyond looks? From similarity-based to cue abstraction processes in multiple-cue judgment. *Developmental Psychology*, *46*(1), 220–229. https://doi.org/10.1037/a0016690

Zhang, Y., Hedo, R., Rivera, A., Rull, R., Richardson, S., & Tu, X. M. (2019). Post hoc power analysis: Is it an informative and meaningful analysis? *General Psychiatry*, *32*(4), Article e100069. https://doi.org/10.1136/gpsych-2019-100069