This postprint was originally published by the American Psychological Association as:
Swaboda, N., Meder, B., & Ruggeri, A. (2022). Finding the (most efficient) way out of a maze is easier than asking (good) questions. *Developmental Psychology, 58*(9), 1730–1746. https://doi.org/10.1037/dev0001386.

Finding the (Most Efficient) Way Out of a Maze Is Easier Than Asking (Good) Questions

Nora Swaboda[1], Björn Meder[1, 2], and Azzurra Ruggeri[1, 3, 4]

[1] MPRG iSearch, Max Planck Institute for Human Development, Berlin, Germany
[2] Department of Health, Health and Medical University Potsdam
[3] School of Education, Technical University Munich
[4] Department of Cognitive Science, Central European University

We investigate whether a spatial representation of a search task supports 4- to 7-year-old children's information-search strategies, relative to their performance in a question-asking game. Children played two computationally and structurally analogous search games: a *spatial search task*, the maze-exploration game, in which they had to discover the path through a maze by removing masks covering its passages; and a *verbal search task*, the 20-questions game, where they had to identify a target monster from a set of eight monsters by asking yes-no questions. Across four experiments, we found that children searched more efficiently when they could make queries nonverbally (Experiments 1 and 2a). We also found that merely providing children with a visual conceptual aid that supports their representation of the hypothesis space (Experiment 2b), or familiarizing them with the hypothesis-space structure (Experiment 3) was not sufficient to improve their search strategies. Together, our results suggest that young children's difficulties in the 20-questions game are mainly driven by the verbal requirements of the task. However, they also demonstrate that efficient search strategies emerge much earlier than previously assumed in tasks that do not rely on verbal question generation. These findings highlight the importance of developing age-appropriate paradigms that capture children's early competence, in order to gain a more comprehensive picture of their emerging information-search abilities.

Keywords: information search, active learning, 20-questions game

Supplementary material to this article is available. For more information see http://hdl.handle.net/21.11116/0000-0006-E9CA-0.

It's Saturday morning and suddenly you remember that today is your godchild's birthday! Surely, any 4-year-old would love a unicorn float! Excited about your idea, you head to the nearest toy store. Once you enter, your spirits drop: How are you going to find what you are looking for within this maze of shelves storing thousands of colorful toys? You decide to explore the corridors yourself. An efficient way to find the unicorn float would be to think about the hierarchical categories it belongs to (e.g., outdoor vs. indoor toys, beach toys, a float) and search the store accordingly (i.e., find the section for outdoor toys, then beach accessories, and then floats). The ability to represent objects as embedded in a hierarchical categorical structure is also a crucial component of asking informative questions. Once you arrive at the party, your godchild curiously tries to guess what the present is ("Is it a book?," "Is it a jigsaw puzzle?"). You cannot help but think that their search would be more effective if they approached the problem top-down, just as you did at the store, by targeting higher-level categories first and narrowing down the space systematically question after question (e.g., "Is it an outdoor toy?," "Is it something for the beach?"). Also, you wonder if they would have been efficient at searching the unicorn float at the store as, in this situation, asking a question would not have been necessary. This article investigates across three studies whether and how a *spatial* representation of a search task supports 4- to 7-year-old children's information-search strategies, relative to their performance in a question-asking game.

The Emergence and Development of Effective Information Search

The development and quality of children's information-search strategies has frequently been investigated using the 20-questions

Nora Swaboda https://orcid.org/0000-0002-1934-305X
Björn Meder https://orcid.org/0000-0002-9326-400X
Azzurra Ruggeri https://orcid.org/0000-0002-0839-1929
Data and study materials are available at https://osf.io/e8gbn/.
We thank Anna-Lina Rauschenbach, Antonia Langenhoff, Mareike Breda, Leon Gellrich, Calvin Paulus, and Carlotta Kreuzer for assistance in recruiting, data collection, and coding. We would also like to thank the Museum für Naturkunde, Labyrinth Kindermuseum and Science Center Spectrum in Berlin for providing research space.
The study was not preregistered.
All illustrations and images were created by the study's authors.
Correspondence concerning this article should be addressed to Nora Swaboda, MPRG iSearch, Max Planck Institute for Human Development, Lentzeallee 94, 14195 Berlin, Germany. Email: swaboda@mpib-berlin.mpg.de

game (e.g., Legare et al., 2013; Meder et al., 2019; Mosher & Hornsby, 1966; Nelson et al., 2014; Ruggeri et al., 2015, 2016; Ruggeri & Feufel, 2015; Ruggeri & Lombrozo, 2014). Its goal is to identify a target object among a given set (e.g., "What present did you get for your godchild?") by asking as few yes-no questions as possible. Initially, the question asker entertains a certain set of hypotheses ("hypothesis space") that includes all the potential solutions to the problem at hand (e.g., all possible items from the toy store). Through a process of elimination, the initial hypothesis space is narrowed down until only one hypothesis, the correct solution, remains. Questions in this game have traditionally been classified into *hypothesis-scanning* questions, which target a single hypothesis within the given set of options (e.g., "Is it the yellow beach ball?"), and *constraint-seeking* questions, which target features shared by several hypotheses (e.g., "Is it an outdoor toy?"). Because constraint-seeking questions are able to rule out multiple hypotheses at each step of the search process (e.g., all indoor toys), they are typically considered more informative than hypothesis-scanning questions (although this is not always true, see Meder et al., 2019; Nelson et al., 2018; Ruggeri et al., 2015; Ruggeri et al., 2017).

Previous research employing the 20-questions paradigm has shown that the informativeness of children's questions steadily increases with age: Until age 7, children almost exclusively ask hypothesis-scanning questions (Herwig, 1982; Mosher & Hornsby, 1966; Ruggeri et al., 2015, 2016; Ruggeri & Feufel, 2015). Children begin to rely more on constraint-seeking questions around age 10, until asking constraint-seeking questions becomes the dominant strategy in adulthood (Ruggeri et al., 2015; Ruggeri & Feufel, 2015). Recently, more computationally-oriented studies, which assess the informativeness of questions using quantitative measures such as expected stepwise *information gain* (e.g., Eimas, 1970), have confirmed this developmental trajectory (Kachergis et al., 2017; Meder et al., 2019; Nelson et al., 2014; Ruggeri et al., 2016; Ruggeri & Feufel, 2015; Ruggeri & Lombrozo, 2014).

There are likely several reasons underlying preschoolers' difficulty with formulating informative questions from scratch (see Ruggeri et al., 2016; Ruggeri et al., 2021). On the one hand, constraint-seeking questions can be more cognitively demanding than hypothesis-scanning ones. For instance, asking an effective constraint-seeking question requires the ability to represent the categories that could be used to partition the presented hypothesis-space at multiple levels of abstraction and to determine which objects belong to the various categories based on certain shared features. Supporting this idea, Ruggeri and Feufel (2015) found that scaffolding more abstract categorical representations facilitated 7- to 10-year-old children's ability to ask constraint-seeking questions. In their study, children and adults were presented with word labels for 20 different objects (e.g., "dog" or "sheep"), describing them either at the basic level (e.g., "dog") or at the subordinate level (e.g., "dalmatian"). The authors found that participants were more likely to ask constraint-seeking questions in the basic-level condition compared to the subordinate condition, suggesting that more abstract labels facilitated a shift away from object-based reasoning when generating questions, allowing them to go beyond the hypothesis-scanning approach. The authors also found that the very ability to generate more abstract features (e.g., "a dog is a mammal" or "a dog has four legs") is one factor that affects performance and that develops across childhood (see also Herwig, 1982). With even younger children, Legare et al. (2013) found that the ability to identify and to flexibly categorize objects based on alternative features (e.g., color and pattern) predicted how well 4- to 6-year-olds could generate effective questions. Also, children perform better in the 20-questions game when they are given the opportunity to *select* among given questions, rather than having to *generate* them from scratch (Herwig, 1982; Ruggeri et al., 2017; Ruggeri & Feufel, 2015), potentially because this eliminates the need to identify and generate those categorical features that can be used to target different levels of the hierarchical structure. Along these lines, Jones et al. (2020) demonstrated that the ability to appreciate the relationship between hypotheses, and therefore to entertain abstract representations of the hypothesis space, develops between age 5 and 8, and can guide information search.

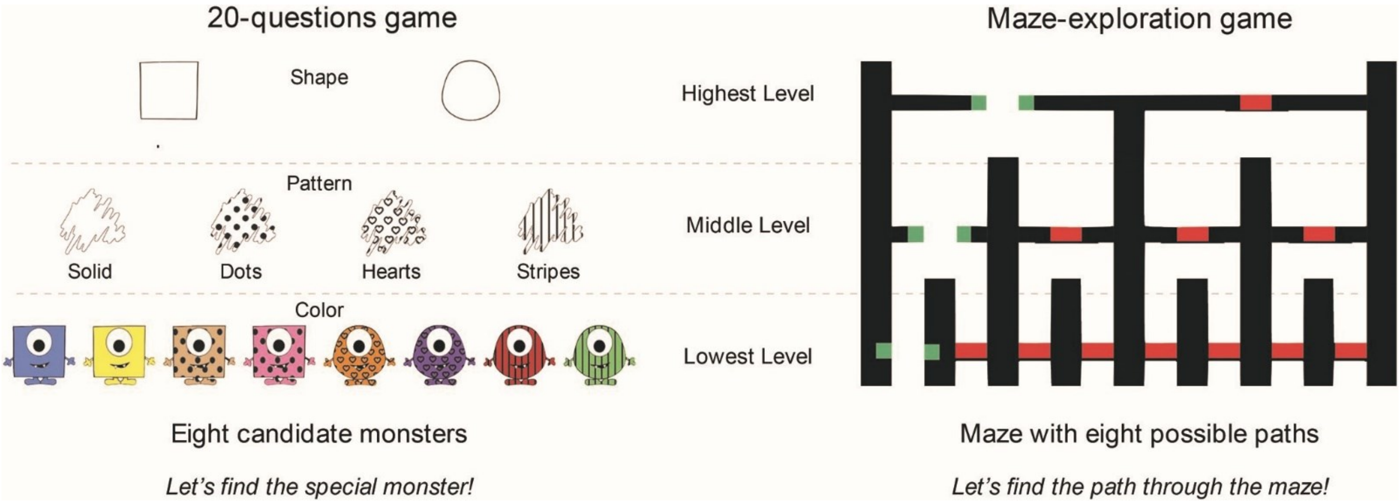
On the other hand, generating constraint-seeking questions may also be more linguistically demanding as it requires children to be familiar with and produce the labels of the different categories and their distinguishing features, and to master various question structures (e.g., "*Is it* an outdoor toy?" vs. "*Does it have* dots?" vs. "*Can it be used for* swimming?"). In contrast, asking hypothesis scanning questions merely requires children to label the candidate objects, which can be accomplished using a single question form (e.g., "Is it the ball?"), or even by simply pointing at the available candidates (i.e., "Is it this one?"). Although the link between verbal skills and performance on the 20-questions game has never been investigated directly, efficient search emerges much earlier in tasks that do not involve verbal question asking (e.g., Bonawitz et al., 2012; Cook et al., 2011; Domberg et al., 2020; Kushnir & Gopnik, 2005; Ruggeri et al., 2019; E. Schulz et al., 2019; L. E. Schulz & Bonawitz, 2007), which suggests a strong involvement of language skills in the development of informative question asking.

Goals and Scope

In this article, we investigate whether 4- to 7-year-old children's information-search strategies can be scaffolded and supported by (a) highlighting the categorical relationships between hypotheses by representing them *spatially*, and (b) reducing verbal demands associated with question generation. In Experiment 1, we presented children with a novel spatial search task, the *maze-exploration* game, in which they had to discover the path through a maze, and a traditional version of the 20-questions game, where they had to identify a target monster from a set of eight monsters by asking yes-no questions. We developed the maze-exploration game to be structurally and computationally analogous to the 20-questions game, so that we could compare children's search performance across both tasks. Just as in the 20-questions game, where the monsters varied across three hierarchically and symmetrically structured features (shape, pattern, and color; Figure 1 right), giving rise to eight unique candidate monsters, the maze was symmetrically organized onto three hierarchical levels, with two passages on the higher, four passages on the middle, and eight passages on the lower level (Figure 1 left). We hypothesized that children would conduct more informative queries in the maze-exploration game than in the 20-questions game because in the spatial search task the structure of the hypothesis space may be more evident and

Figure 1

Illustration of the Structural Equivalence Between the 20-Questions and Maze-Exploration Game



Note. In the 20-questions game (left), the eight candidate monsters were made up of three hierarchically organized features (shape, pattern, color; assignment of features to each level counterbalanced across participants). The maze (right) also consisted of three hierarchically organized levels (higher, middle, lower). To ensure that the tasks were structurally and computationally analogous, there was only one correct path through the maze and only one open passage on each level. See the online article for the color version of this figure.

because searching would not require children to ask verbal questions.

Experiment 1

Method

Participants

Participants were 124 children between 4 and 7 years (63 girls, 61 boys, $M = 73.6$ months, $SD = 14.59$ months), recruited at local museums and preschools in Berlin, Germany. An additional 21 children were excluded from analysis (seven due to experimenter error, two due to insufficient command of German, and 12 because they did not complete the task. Of the 12 children who did not complete the task, ten failed to ask questions in the 20-questions game (three 4-year-olds, two 5-year-olds, and five 6-year-olds) and two failed to understand the rules of the maze-exploration game (both were 4-year-olds). Written informed consent of legal guardians was obtained prior to participation. Children were asked for verbal consent and received stickers as a reward for their participation (see Design and Procedure section). Sample size was determined by conducting a-priori power calculations for the planned statistical tests. The most conservative estimate indicates an overall sample of 88 children to detect an estimated effect size of Cohen's $d = .20$ with 90% power and a .05 criterion for statistical significance for a mixed model linear regression with 4 predictors (task, age, Task × Age, and participant). Power analyses were performed using G*Power V3.1.9.6 (Faul et al., 2007). Additional participants were collected in order to even out the age distribution across the different age groups. All reported studies (Experiments 1, 2, and 3) were approved by the Ethics Committee of the Max Planck Institute for Human Development in Berlin (study name: Maze; protocol: MAZE). Study materials, analysis code and data are available on the project's OSF repository. This study was not preregistered.

Design and Procedure

Children performed two tasks (order counterbalanced): a 20-questions game and a maze-exploration game. In the 20-questions game, the goal was to find out which out of eight monsters activated a music machine by asking as few yes-no questions as possible. In the maze-exploration game, children had to discover the path leading through a partially blocked two-dimensional maze by finding out which of its hidden passages were open. Children could then move a marble through the maze along the discovered path to "charge" it, giving it the power to activate the music machine. The two tasks were designed to be structurally and computationally analogous.

20-Questions Game

At the beginning of the game, the experimenter presented children with a set of eight monster cards and demonstrated that one monster (the target monster) activated a music machine when inserted into it (without showing which monster it was). The eight monsters varied across three hierarchically structured features (shape, pattern, and color), such that four of the eight monsters shared the higher-level feature (two variants; for example, square and circle), two monsters shared the middle-level feature (four variants; for example, solid, dots, hearts, and stripes), and each monster was unique with respect to the lower-level feature (eight variants; for example, blue, yellow, brown, pink, orange, purple, red, and green, see Figure 1 left). We counterbalanced the level to which each feature was assigned across six sets of monsters (see online supplementary materials). The monster set and target monster were assigned to children pseudorandomly. The experimenter then shuffled the monster cards and arranged them in a 2 × 4 grid on a table, visually organized according to their hierarchical structure (for example, square monsters on the left, round monsters on the right; solid monsters on top, dotted monsters on the bottom, see Figure 2 left). Children were instructed to find the target monster by asking as few yes-no questions as possible. Children could ask as many questions as they wanted, but had to pay a sticker for every question, out of a set of eight stickers they were initially given. They could take home all remaining stickers at the end of the game. Following each question, the experimenter moved all eliminated monsters (that is, those inconsistent with the obtained information) on a "no-music" pile. This procedure was repeated until the child had identified the target monster, that is, when only one monster remained. The child could then insert the target monster into the music machine to activate it.

Maze-Exploration Game

At the beginning of the game, the experimenter demonstrated how a small marble could activate the music machine. She then pointed to the maze, whose passages were concealed with yellow cards (Figure 2 right), and explained that once the marble activated the machine, it lost its special powers, but that it could get recharged by going through the maze. The maze was symmetrically organized onto three hierarchical levels, with two passages at the higher, four passages at the middle, and eight passages at the lower level (see Figure 1). To achieve structural and computational equivalence with the 20-questions game, there was only one correct path through the maze and only one open passage per level (that is, no dead ends). Consequently, there were exactly eight possible maze configurations (that is, eight potential paths through the maze, each leading to one of the eight lower-level passages), to which children were assigned pseudorandomly. To rule out a potential bias to start from the top (for example, due to expectations about gravity), the maze was rotated 90°, so that the hierarchical levels were organized right-to-left or left-to-right (orientation counterbalanced). For simplicity and consistency with the 20-questions game, we will refer to the levels according to their default, unrotated hierarchical position as "higher" (two passages), "middle" (four passages), and "lower" (eight passages) levels.

To familiarize children with the structure of the maze and the constraints (for example, exactly one open passage per level), children were then presented with pictures illustrating all eight possible fully-revealed maze configurations (presented in random order, and in the same rotated orientation as the real maze) and were asked to indicate the path through each of them by tracing it with their fingers. Subsequently, the experimenter pointed to the real maze and explained that, to find the path through this particular maze, children would have to find out which of the hidden passages were open or closed, and that they could do so by removing the yellow cards that were covering them. As in the 20-questions game, children were allowed to make as many queries as they wanted, but were incentivized to find the path with as few queries as possible by having to pay a sticker (from an initial set of eight stickers) for each query. Once children had revealed a passage, the experimenter provided feedback by pointing out whether the discovered passage was open or closed. In addition, just as she removed the eliminated monsters from the display in the 20-questions game, she uncovered those passages of the maze children could make inferences about due to implicit horizontal and vertical dependencies. For instance, if a child discovered that the right passage was open on the higher level, the experimenter revealed the remaining closed passage on the left, as well as all closed passages on the middle and lower level below it. This is equivalent to discovering that the target monster is a square and removing all four round monsters from display (for the full set of dependencies, see online supplementary materials). After children had found the correct path, the experimenter placed the silver marble at the entrance of the maze, so that children could move it across and use it to turn on the music machine.

Results

Mean Information Gain of Participants' Queries

We calculated the expected stepwise information gain (see online supplementary materials for details) of each participant's queries in the two tasks (Figure 3a). We then analyzed the informativeness of participants' queries with a mixed-model linear regression with task (20-questions game vs. maze-exploration game), age (in months), and their interaction as predictors, and expected information gain as dependent variable. As every participant provided multiple observations, we added varying intercepts for each participant. The model revealed that task, age and the interaction of these two factors were significant predictors of children's mean expected information gain (see online supplementary materials, Table S2): Children made more informative queries in the maze-exploration game than in the 20-questions game, and older children made more informative queries than younger ones, with the difference between tasks decreasing with age (see Figure 3a).

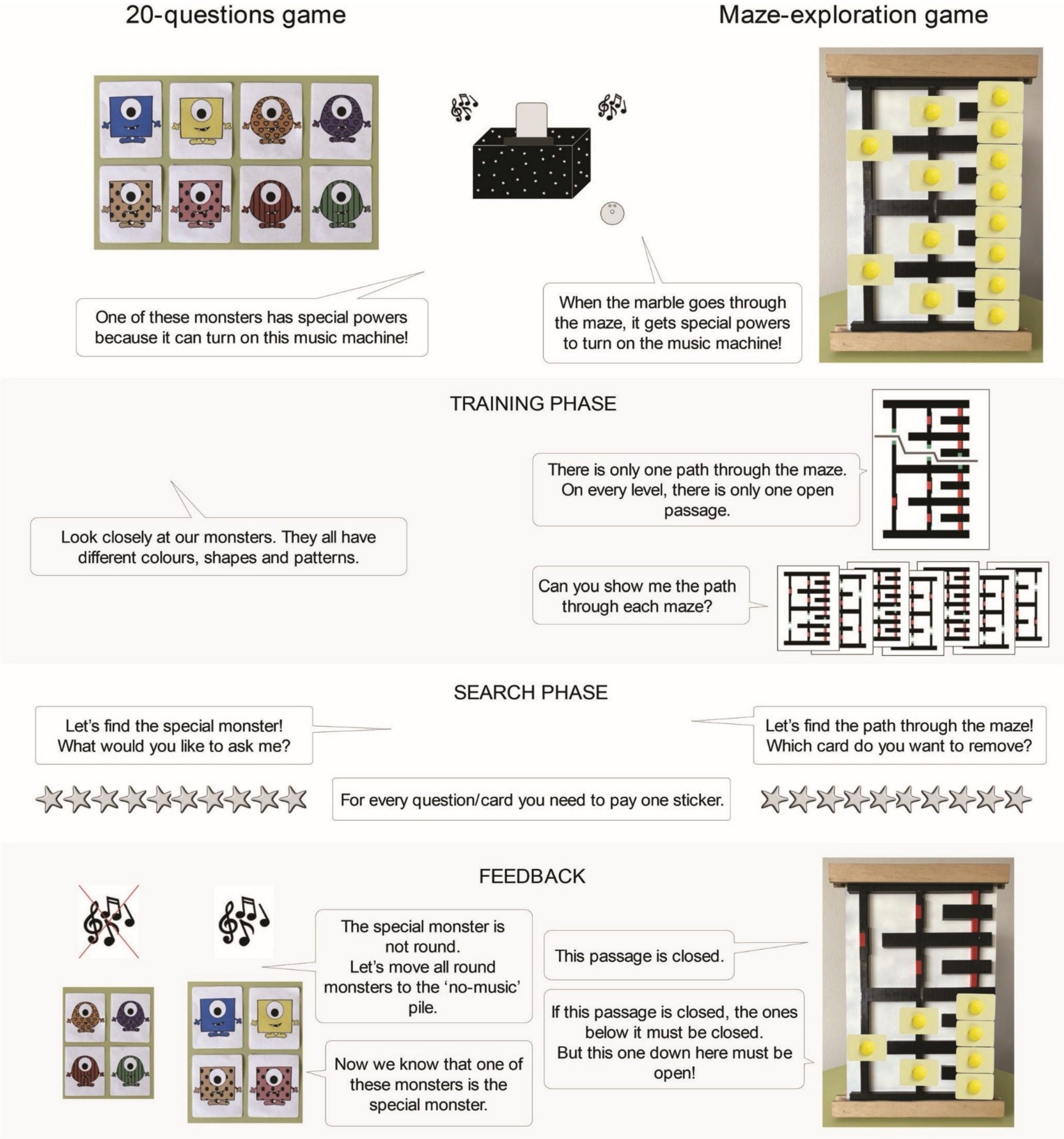
In a separate regression model with the predictors task (20-questions game vs. maze-exploration game), order (first vs. second) and their interaction, we also looked at whether the order in which tasks were presented had an effect on the informativeness of children's queries. This analysis confirmed a main effect of task, with performance higher in the maze-exploration game than in the 20-question game, and a significant interaction between task and order (Table S2). Post hoc t-tests revealed that performance on the 20-questions game was not affected by whether children played the game first or second, $t(120) = 1.60$, $p = .113$, whereas performance was higher in the maze-exploration game when children played it second, as compared with first, $t(118) = -2.25$, $p = .03$.

Children's First Queries

Figure 3b shows the proportion of children targeting each level with their first query in the two tasks. For visualization purposes, age is broken down into groups (4-year-olds, 5-year-olds, 6-year-olds, 7-year-olds). To analyze these proportions, we looked at whether or not children performed the most informative first query (that is, targeted the highest level of the hierarchy, see Figure 4a) and performed a mixed-model binary logistic regression with level of first query (1 = higher level/most informative, 0 = other levels) as dependent variable and task (20-questions game versus mazeexploration game) and age (in months) as predictors. We also added varying intercepts for each participant as there were multiple observations per participant. Both task and age group were significant predictors of whether children performed the most informative first query (online supplementary materials, Table S3): children had an increased likelihood of targeting the higher

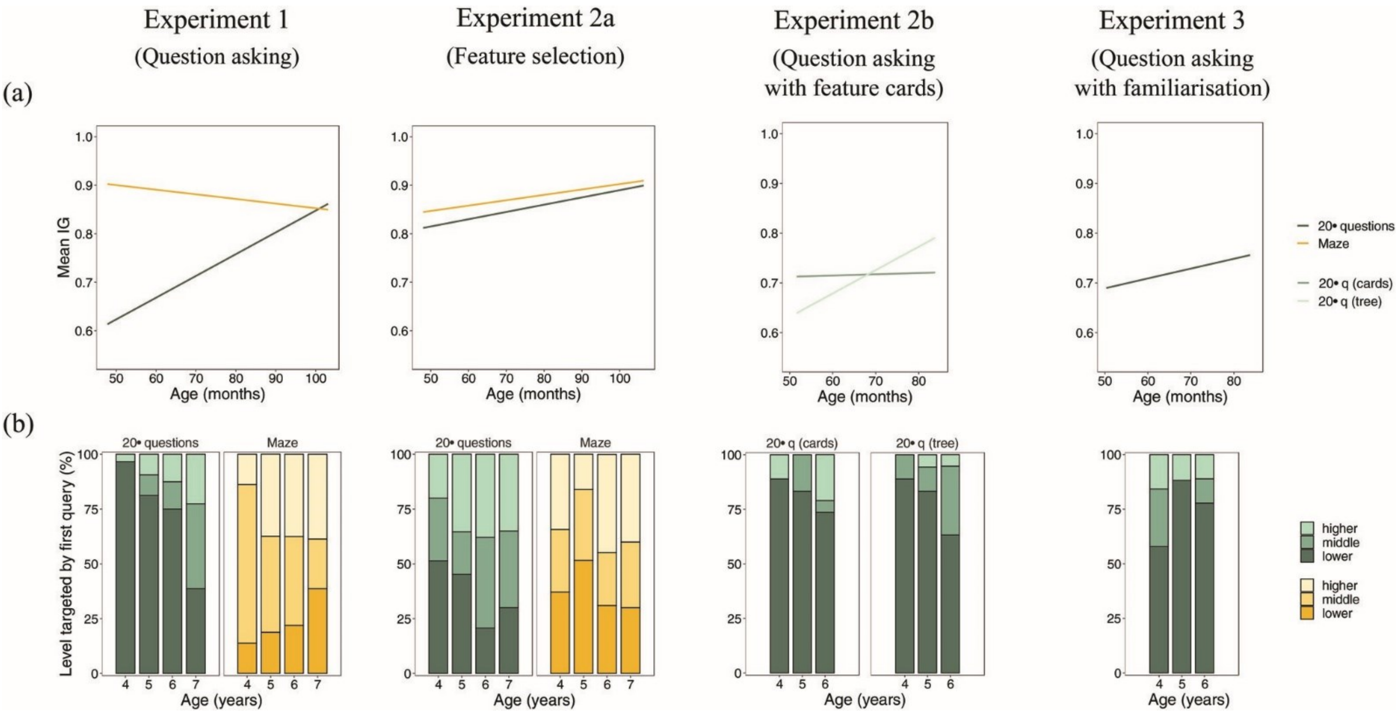
Figure 2

Procedure of Experiment 1



Note. Children played a 20-questions game, where they had to identify the target monster from a set of eight monsters (left), and a maze-exploration game, where they had to find the path through a maze (right). To familiarize children with the structure of the maze, they saw an image of every possible configuration and indicated the path leading through it. Children made their queries by asking yes-no questions and by removing the yellow cards hiding the maze's passages. They paid one sticker for every query. Feedback was provided after every query by moving eliminated monsters to a "no-music" pile, and by uncovering all eliminated passages. See the online article for the color version of this figure.

Figure 3
Informativeness of Children's Queries



Note. In Experiment 1, children made more informative queries in the maze-exploration game than in 20-questions game (a), and were more likely to target the highest level of the hierarchy, that is, to make the most informative first query (b). In Experiment 2a, when allowed to use feature cards as response aid, children made equally informative queries (a), and were equally likely to target the highest level of the hierarchy (b) in both tasks. In Experiment 2b, where children were provided a visual display of the hypothesis space but had to ask verbal questions, and Experiment 3, where they completed a training phase akin to the training in the maze-exploration game, the informativeness of children's questions dropped to the level observed in Experiment 1. Dots (a) represent individual participants, solid line is fitted regression line, and shaded regions indicate 95% confidence intervals. Color shades in bar charts (b) represent the three hierarchical levels (dark = lower, medium = middle, light = higher). See the online article for the color version of this figure.

level of the hierarchy in the maze compared with the 20-questions game, and older children were more likely to do so than younger children (see Figure 3b).

Proportion of Children Employing the Optimal Strategy

In both tasks, the optimal strategy consisted of a “top-down” approach, sequentially targeting the higher, middle, and lower level. We classified children according to whether they implemented the optimal strategy, therefore achieving a mean information gain of 1, and those who did not. Figure 4b shows the proportion of children employing the optimal strategy in the 20-questions versus the maze-exploration game.[1] We conducted a mixed model logistic regression with implementation of the optimal strategy as dependent variable, and task (20-questions game versus maze-exploration game) and age group (in months) as predictors. The analysis revealed that both task and age were significant predictors (see online supplementary materials, Table S4): children were more likely to implement the optimal strategy in the maze-exploration than in the 20-questions game, and older children were more likely to do so than younger ones (see Figure 4b).

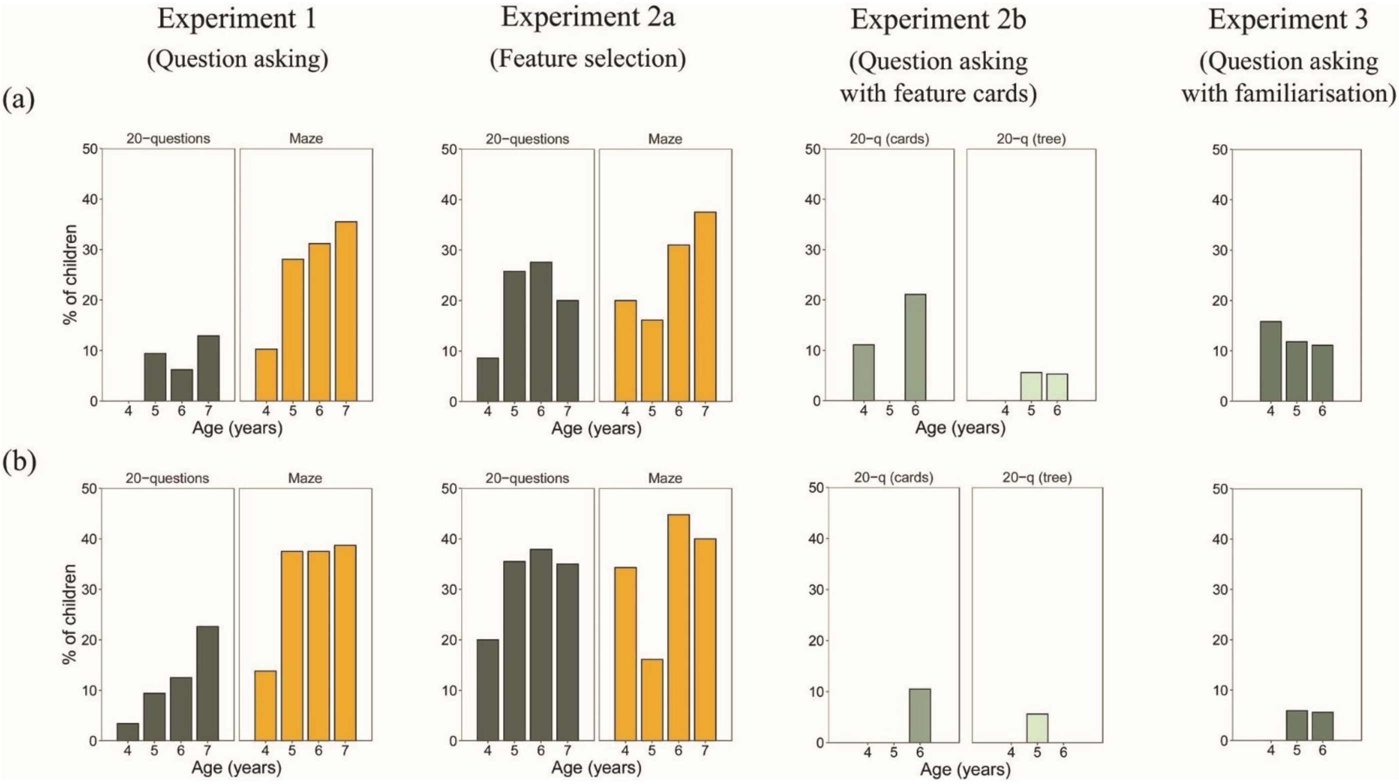
Discussion

The results of Experiment 1 show that children employed more informative search strategies in the maze-exploration game than in the 20-questions game: they achieved higher mean information gain, and were more likely to make the most informative first query and to employ the overall optimal strategy. Indeed, the maze was so effective that it boosted younger children’s search performance up to the level achieved by 7-year-olds. This result indicates that, under the right conditions, the ability to conduct informative queries may emerge at an earlier age than previously assumed. In contrast, and in line with previous research, the quality of children’s questions in the 20-questions game increased with age (Herwig, 1982; Mosher & Hornsby, 1966; Ruggeri et al., 2015, 2016; Ruggeri & Feufel, 2015). In this sense, younger children especially benefited from the maze, with performance differences between tasks decreasing with age and disappearing by age 7. This interestingly corresponds to one of the milestones in the development of children’s information-search strategies (Jones et al., 2020), the point when children begin to ask more informative constraint-seeking questions (Herwig, 1982; Mosher & Hornsby, 1966; Ruggeri et al., 2015, 2016; Ruggeri & Feufel, 2015). On the one hand, because performance of 7-year-olds in our 20-questions game was already fairly high, there might not have been enough room to observe a beneficial effect of the spatial layout. On the

[1] Note that these numbers reflect children who employed the optimal strategy in one *or* the other task, or in both. Only four children (two 6-year-olds and two 7-year-olds) followed the optimal strategy in both games.

Figure 4

Percentage of Children Who Made the Optimal First Query (a) and Who Employed the Optimal Overall Strategy (b) in the Four Experiments, Broken Down Into Age Groups for Visualization



Note. In Experiment 1, children were significantly more likely to make the most informative first query (a), and to employ the optimal overall strategy (b), in the maze-exploration game than in the 20-questions game. Older children were more likely to employ the optimal first query, and overall optimal strategy than younger ones. There were no differences between tasks in Experiment 2a, nor between versions in Experiment 2b. See the online article for the color version of this figure.

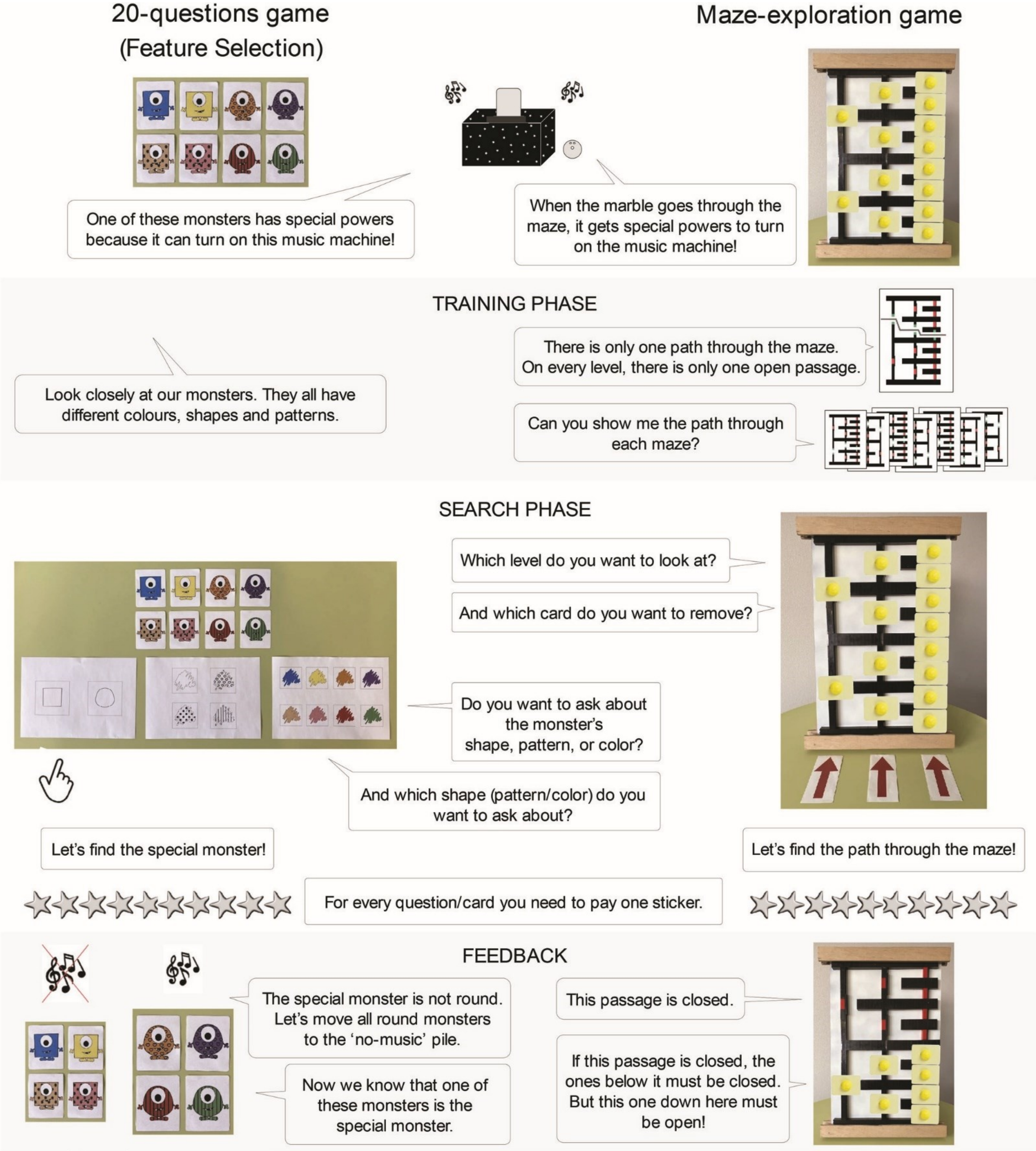
other hand, our findings may indicate that the support offered by the spatial layout of the maze-exploration game might no longer be helpful after a certain stage of development.

Our results also indicate that performance on the maze-exploration game particularly improved when children played it *after* the 20-questions game (as compared with before), generally suggesting that there is margin for individual improvement on the maze-exploration task. Our current design does not allow us to determine whether this effect was specific to playing the 20-questions game first (for example, preexposure to the monsters may have facilitated children's thinking about hierarchical structures, allowing them to exploit the spatial structure of the maze more efficiently), or whether it arose more generally from having engaged in another active learning task, or just from being more familiar with the experimenter. In reverse, there was no training effect from the maze-exploration game to the 20-questions game, possibly because training effects on this task are generally very hard to achieve (for example, Courage, 1989; D. R. Denney, 1972; D. R. Denney et al., 1973; N. W. Denney & Turner, 1979), or because performance is too much constrained by general cognitive factors, and especially verbal competence, for such an effect to manifest. Future work may further investigate how to scaffold and train active learning competence across different tasks.

The spatial search task could have supported younger children's search strategies in two ways: by highlighting the hierarchical, nested structure of the hypothesis space—an aspect that is known to develop across these childhood years (Jones et al., 2020), or by eliminating verbal requirements from the search (Ruggeri et al., 2017). To disentangle these alternative explanations, we conducted a follow-up study (Experiment 2a), in which we compared children's performance in the same maze-exploration game, and in a novel nonverbal *feature-selection* version of the traditional 20-questions game. In this version, children were presented with three cards illustrating the different monster features (that is, shape, color, and pattern, see Figure 5 left) and could point at the feature they wanted to query, similarly to pointing at the passage in the maze they wanted to reveal. If the performance boost observed in the spatial search task in Experiment 1 was driven by the fact that it did not rely on their verbal skills, younger children in Experiment 2a should achieve the same informativeness in both games. In contrast, if the maze-exploration game was beneficial because its spatial layout highlights the hierarchical structure of the hypothesis space, then task-dependent differences should persist.

In addition, the difference in performance between our two tasks may also be attributable to children completing an extensive *training phase* familiarizing them with the task structure in the maze-exploration game, but not in the 20-questions game. To rule

Figure 5
Procedure of Experiment 2a



Note. Children played a 20-questions game (left) and a maze-exploration game (right). In the 20-questions game, three cards illustrating the monster features served as response aid. In the maze, three analogous red arrows indicated the three levels. Children made their queries by first selecting (pointing at or naming) the feature (maze level), and then the feature variant (maze passage) they wished to query. They paid one sticker for every query. Feedback was provided after every query by moving eliminated monsters to a "no-music" pile, and by uncovering all eliminated passages. See the online article for the color version of this figure.

out this alternative explanation, we conducted a second follow-up study (Experiment 3): children played a 20-questions game, but were familiarized with the monsters and their features in the same way as they had been familiarized with the structure of the maze in Experiments 1 and 2a: picking one random monster, the experimenter explained that each monster was made of three features (e.g., “This monster is a square, is red, and has stripes.”). Children then labeled the three features of the remaining seven candidate monsters. If, with training, children’s performance in Experiment 3 resembles performance in the original maze-exploration game, the beneficial effects we observed in Experiment 1 likely arose because of the additional training children received in the maze-exploration game. If children in Experiment 3 continue to perform worse in the 20-questions game than in the original maze-exploration game, the difference in training likely did not contribute to performance differences between the two tasks in Experiment 1.

Experiment 2a

Method

Participants

Participants in Experiment 2a were 135 children between 4 and 7 years (66 girls, 67 boys, $M = 73.82$ months, $SD = 16.20$ months), recruited at the same local museums as children in Experiment 1. An additional 11 participants were excluded due to experimenter error ($n = 6$) and video failure ($n = 5$). As in Experiment 1, sample size was determined by conducting a-priori power calculations for the planned statistical tests. This produced an estimate of 88 children for a mixed model linear regression with four predictors (task, age, Task × Age, and participant) with an estimated effect size of Cohen’s $d = .20$, 90% power and a .05 criterion for statistical significance. Additional participants were collected in order to even out the age distribution across the tested age groups.

Materials, Design, and Procedure

Materials and procedure were largely identical to those of Experiment 1, with two differences. First, in the feature-selection game, children were presented with three *feature cards*, each depicting all variants of one feature (that is, shape, color, and pattern; see Figure 5 left). These cards served as a response aid for children, who could use them to select (that is, point at) the feature and variant they wished to query, instead of having to verbalize a question. As in Experiment 1, children were instructed to ask questions to find the target monster, that is, they were not explicitly prompted to use the cards. However, they could refer to them *if they wanted to*. Second, the question-asking procedure was broken down into two steps, where the experimenter, referring to the feature cards, first asked children to select the feature they wanted to query (“What do you want to ask about first (next)—the monster’s shape, pattern or color?”). Only after children had made their choice, the experimenter followed up to clarify the variant, saying “And which shape (color/pattern) do you want to ask about?”). If children asked about/pointed at a feature value (or a candidate monster) right away, no follow-up question was asked. The monsters ruled out by a child’s question, as well as the feature cards for which the target variant had been identified, were removed from display. The feature cards were presented in their hierarchical order matching the orientation of the maze. Hence, from left to right, the feature cards were either presented in the order “lower–middle–higher” level or “higher–middle–lower” level (order counterbalanced).

To ensure an analogous procedure in the maze-exploration game, we introduced three red arrows pointing at the three different levels (Figure 5 right). When making a query, children were instructed to first select the arrow pointing at the level they wished to query, before selecting the passage on the chosen level they wished to reveal. As in the 20-questions game, if children pointed at a passage right away (i.e., did not use the arrows), the experimenter accepted their selection without additional prompts. Arrows were removed from display once the open passage on that level had been found.

Results

Mean Information Gain of Participants’ Queries

Figure 3a shows the information gain of participants’ queries in Experiment 2a. A mixed-model linear regression with the predictors task (feature-selection game vs. maze-exploration game) and age (in months), and mean information gain as dependent variable revealed that age, but not task, was a significant predictor of mean information gain, with older children making more informative queries than younger ones (online supplementary materials, Table S2).[2]

Proportion of Children Employing the Optimal Strategy

Figure 4b shows the proportion of children employing the optimal strategy in the two tasks broken down by age (in years). A mixed-model logistic regression with implementation of the optimal strategy dummy coded as 1 (optimal strategy) and 0 (nonoptimal strategy), and task (feature-selection game vs. maze-exploration game) and age (in months) as predictors revealed that children were equally likely to employ the overall optimal strategy in both tasks, and that younger children were as likely as older ones to do so (online supplementary materials, Table S4).

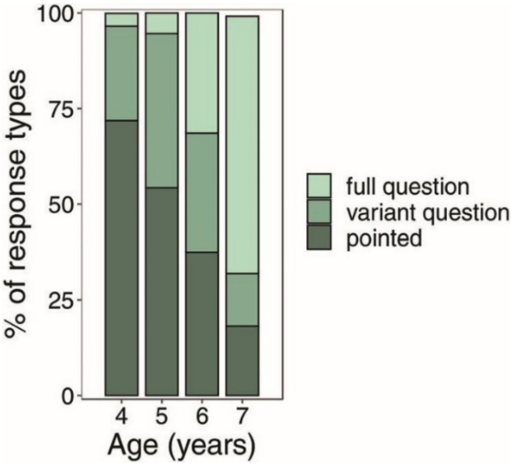
Proportion of “Pointing” Queries in the 20-Questions Game

In Experiment 2a, children made their queries in three different ways (“response types”): by pointing at the feature cards or candidate monsters (“pointing”), by naming the feature variant they wished to query without embedding it in a sentence (“variant question”; e.g., “Blue?”), or by asking a full verbal question (“full question”; e.g., “Is the special monster blue?”).[3] We first computed the relative frequency of the three response types for each child. Figure 6 shows the proportion of children’s questions made using the three response types, binned into age groups for visualization (4-, 5-, 6-, and 7-year-olds). Three separate linear regressions (one for every response type) revealed that age was a significant predictor of the relative frequency of these response

[2] Note that including an interaction term did not improve model fit.

[3] Note that children did not necessarily use the same type of response for all their questions.

Figure 6
Relative Frequency of the Three Response Types in the 20-Questions Game (Experiment 2a)



Note. Children made their queries either by pointing at the feature cards or candidate monsters ("pointing"), by naming the feature variant they wished to query without embedding it in a sentence ("variant question," e.g., "Blue?"), or by asking a full verbal question ("full question," e.g., "Is the special monster blue?"). Age has been binned into separate age groups for visualization purposes. See the online article for the color version of this figure.

types, with the proportion of children's questions made by pointing at the cards decreasing with age and the proportion of variant questions and full questions increasing with age (online supplementary materials, Table S5).

Proportion of Children Using Pointing to Make Queries

We performed two binary logistic regressions to analyze whether age (in months) predicted how likely children were to use pointing for all their queries (dummy coded as 1 = always pointed, 0 = pointed sometimes), or at least once (1 = pointed for one or more queries, 0 = did not point at all). The analyses revealed that age was a significant predictor of both (online supplementary materials, Table S6), with younger children significantly more likely to point exclusively than older children (54.3% of 4-year-olds, 35.5% of 5-year-olds, 27.6% of 6-year-olds, and 10.0% of 7-year-olds), and significantly more likely to point at least once as compared to older children (82.9% of 4-year-olds, 74.2% of 5-year-olds, 55.2% of 6-year-olds, and 30.0% of 7-year-olds).

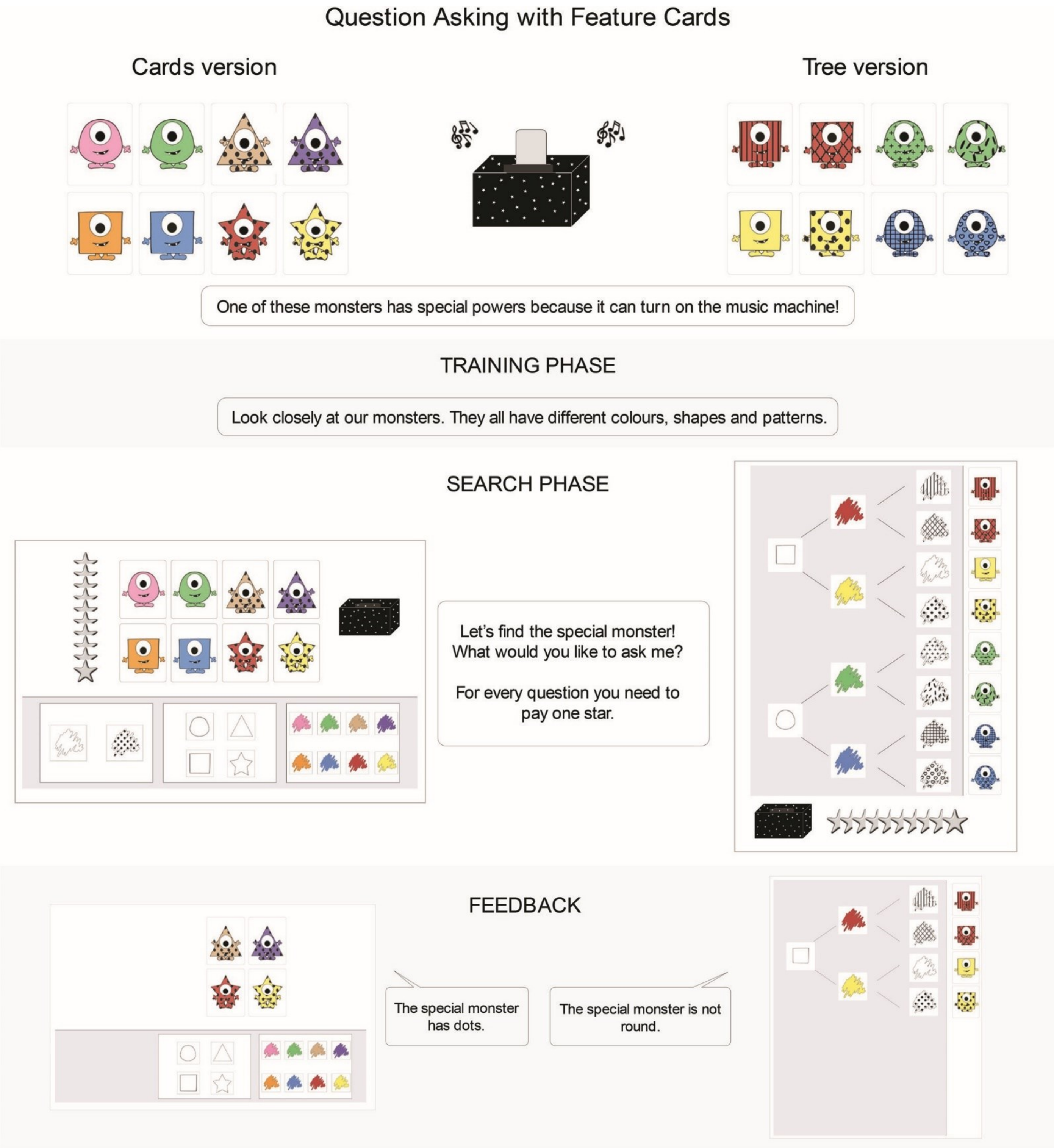
Discussion

Experiment 2a examined to what extent the beneficial effect of the maze-exploration game observed in Experiment 1 can be attributed to the fact that children did not need to verbalize questions, but could make queries by pointing at the maze passages. To do so, we introduced three feature cards to the 20-questions game, each depicting all possible variants of one monster feature (e.g., plain and dotted for pattern; circle, square, triangle, and star for shape; and pink, green, orange, blue, brown, purple, red, and yellow for color, see Figure 5). Children could use these cards as a *response aid* by pointing at the feature and variant they wished to query (akin to pointing at the maze passages). The results of Experiment 2a show that children achieved the same performance in the feature-selection version of the 20-questions game as in the maze-exploration game, and that older children made more informative queries than younger ones, although the age benefit was less pronounced than in Experiment 1. The analysis of *how* children made their queries revealed that, compared to older children, younger children were more likely to point at the cards suggesting that the response aid was particularly beneficial for the younger children. Furthermore, our analysis showed that in transitioning from pointing to asking complete verbal questions, which was the predominant response type for older children from about age 7, children passed through an intermediate strategy of naming the feature variants, but relying on the experimenter to provide the sentence structure for their question (e.g., "Which color do you want to ask about?," "Blue!"). This strongly suggests that improving language skills, and more particularly syntactical skills, can, at least partially, account for developmental improvements in question asking.

On the other hand, our feature cards may also have served as a *conceptual aid*. In particular, they may have enhanced children's search efficiency by breaking down the hypothesis space into its different categorical features and illustrating the number of values each feature could take on (similarly to the spatial maze structure). This might have helped children in identifying which features would be most informative to query. In addition, use of the feature cards was accompanied by a scaffolding process, which broke down each question into two steps (first selecting a feature, and then a variant to query). This process effectively obligated children to think about the different categories (i.e., shape, color, or pattern) before picking a variant, which may have additionally contributed to the cards' role as conceptual aid. Indeed, a recent study has shown that reminding children of the categories that can be used to ask (informative) questions can improve their question-asking efficiency (Ruggeri et al., 2021): In a similar 20-questions game design, children had to identify a target monster from a set of monsters differing in color, shape, and pattern. Before asking a question, the experimenter verbally reminded children of the features they could query ("It could be that the monsters with a specific pattern—dotted or not dotted—turn on the machine. Or, it could be that monsters with a specific body shape—square, round, spiked, or triangular—turn on the machine. Or, it could be that monsters of a specific color—orange, purple, white, pink, green, red, yellow, blue—turn on the machine"). It is therefore possible that children in Experiment 2a benefited from the feature cards because they provided a conceptual aid.

Experiment 2b addresses this confound, examining whether the feature cards in Experiment 2a served as response aid or rather as conceptual aid, or both. Children played two versions of the 20-questions game (order counterbalanced), where we provided them with a spatial display of the categories constituting the hypothesis space, but removed the possibility of using this display as a response aid (see Figure 7 left). We varied how the hypothesis-space was presented: in the *cards version*, children were given three cards illustrating the values each feature could take on (as in Experiment 2a). In the *tree version*, the hypothesis space was presented as a hierarchical tree structure (i.e., using the same spatial organization as in the maze-exploration game, see Figure 7 right).

Figure 7
Procedure of Experiment 2b



Note. Children played two 20-questions games (order counterbalanced): In the cards version, they were given the monster features on separate cards, as in Experiment 2a. In the tree version, the features were presented spatially in their hierarchical tree structure (akin to the maze in Experiment 1). They had to ask yes-no questions to find the monster, and paid one sticker for every question asked. Sessions were conducted via an online videoconferencing platform. See the online article for the color version of this figure.

Unlike in Experiment 2a, children had to ask *verbal* questions to identify the target monster, and could no longer point at the cards to make their queries. Indeed, pointing was prevented by the online testing mode of Experiment 2b. If the feature cards serve as conceptual aid, then children's performance on this task should resemble the performance observed in the feature-selection version of the 20-questions game (i.e., Experiment 2a). If, in contrast or in addition, the cards served as response aid, children's performance should be worse compared with Experiment 2a.

Experiment 2b

Method

Participants

Participants in Experiment 2b were 46 children between 4 and 6 years (19 girls, 27 boys, $M = 69.33$ months, $SD = 9.09$ months). They were recruited via an online database and tested virtually using the video-conferencing platform BigBlueButton. An additional 11 participants were excluded due to experimenter error ($n = 7$), parental interference ($n = 2$), or because of insufficient command of German (non-natives, $n = 2$). a priori power calculations for Experiment 2b produced a sample size estimate of 52 children for a mixed model linear regression with two predictors (version and participant) with an estimated effect size of Cohen's $d = .20$, 80% power and a .05 criterion for statistical significance.

Design and Procedure

Experiment 2b aimed to investigate the extent to which the feature cards in Experiment 2a served as conceptual aid rather than as response aid. To do so, it combined aspects of Experiment 1 and Experiment 2a: Children played the 20-questions game and had to ask *verbal* yes/no-questions (without scaffolding of the question-asking process) to identify the target monster (as in Experiment 1). They were also presented with a display breaking down the hypothesis-space into its different categories (as in Experiment 2a). However, children could not use this display to make their queries (i.e., they could not point at it, see below) so that it could serve as conceptual, but not as response aid. We varied how this conceptual aid was displayed: In the *cards version*, each feature was represented on separate cards (identical to Experiment 2a, see Figure 7 left). In the *tree version*, features were broken down into the hierarchical tree structure (akin to the maze structure, see Figure 7 right). Order of the two versions was counterbalanced. A different monster set was used for the two conditions. To avoid potential carry-over effects, each set pair consisted of two monster sets that did not share any feature on the same level (see online supplementary materials for all set pairs used). As in Experiment 2a, the features were presented either from higher-to-lower or from lower-to-higher level (counterbalanced across participants, but constant within each participant).

Due to COVID-19-related social distancing regulations in place at the time of testing, Experiment 2b was conducted online. Materials and procedure therefore had to be adapted slightly for online testing. Children saw the stimuli on their own computer screen at home and an experimenter guided them through the tasks: The experimenter first introduced the music machine and explained that one special monster could activate it. At the same time, an animation showed that the music machine played music after a gray card was inserted into it. The card was then moved into a deck of cards (akin to shuffling the physical cards in Studies 1 and 2a). Next, the eight target monsters appeared on screen, hierarchically sorted into a 2 × 4 grid, but appearing in random order. Children were then told that they had to find the target monster, and could do so by asking yes-no questions. They could ask as many questions as they wanted, but had to pay one of eight stars (instead of stickers) for every question asked. Next, the feature cards appeared on screen and were introduced by the experimenter saying, "Here, you can see which patterns, shapes, or colors, the monsters can have." In the cards version, the feature cards were placed below the 2 × 4 grid. In the tree version, the monsters were first shifted around to form a straight line, next to which the hierarchical tree structure appeared (see Figure 7 right).

Children were then prompted to ask questions ("What do you want to ask about first (next)?" The monsters ruled out by a child's question, as well as the features for which the target variant had been identified, were removed from display. Crucially, unlike in Experiment 2a, children in Experiment 2b had to make their queries by asking direct verbal questions. Pointing at the cards was made impossible by the online mode of testing, as the experimenter could not to see where the child was pointing. Hence, the feature displays could serve as conceptual aid, illustrating the hierarchical structure of the hypothesis space, but not as response aid.

Results

Mean Information Gain of Children's Queries

Figure 3a shows the mean information gain of participants' queries in Experiment 2b. A mixed-model linear regression with the predictors version (cards vs. tree) and age (in months), and mean information gain as dependent variable showed no significant effect of age, nor of version (online supplementary materials, Table S2).

Children's First Queries and Overall Strategy Optimality

Figure 3b shows the proportion of children targeting each level with their first query in the cards and the tree version of the 20-questions game in Experiment 2b. Overall, 9.1% of children made the optimal first query in the cards version (one 4-year-old, three 6-year-olds) and 4.5% of children in the tree version (one 5-year-old, and one 6-year-old, see Figure 4a). The number of children employing the optimal overall search strategy was also extremely low—one 5-year-old in the tree version, and one 6-year-old in the cards version (see Figure 4b).

Cross-Experiment Comparison of Informativeness of Children's Questions in the 20-Questions Game

To assess whether children in Experiment 2a benefited of the feature displays because these served as response aid, or as conceptual aid, we compared the informativeness of children's questions in the different versions of the 20-questions game in Experiment 1, Experiment 2a, and Experiment 2b. To do so, we built two linear regression models (one for the cards version, and one for the tree version) with expected information gain as dependent variable and study (Experiment 1 vs. Experiment 2a vs. Experiment 2b) and age (in months)[4] as predictors. Both models showed the same pattern of results, with children achieving similar levels of performance in Experiments 2b and Experiment 1, but making significantly more informative queries in Experiment 2a (feature selection) than in Experiment 2b. Age was also a significant predictor of the informativeness of children's queries (online supplementary materials, Table S7).

Experiment 3

Method

Participants

Participants in Experiment 3 were 54 children between 4 and 6 years (23 girls, 31 boys, $M = 65.33$ months, $SD = 10.28$ months). They were recruited via an online database and tested virtually using the video-conferencing platform BigBlueButton. An additional seven participants were excluded due to experimenter error ($n = 2$), parental interference ($n = 4$), or because of insufficient command of German (non-natives, $n = 1$). a priori power calculations for Experiment 3 produced a sample size estimate of 42 children for a mixed model linear regression with one predictor (age) with an estimated effect size of Cohen's $d = .20$, 80% power and a .05 criterion for statistical significance.

Materials, Design, and Procedure

Experiment 3 assessed to what extent difference in training children received in the maze-exploration game and the 20-questions game (i.e., familiarization for the maze, but not for the monsters) may account for children's higher performance on the maze-exploration game in Experiment 1. Children played one 20-questions game but were familiarized with the monsters and their features in the same way as they were familiarized with the structure of the maze in Experiments 1 and 2a: Picking one random monster, the experimenter explained that each monster was made of three features (e.g., "This monster is a square, is red, and has stripes."). The corresponding image was highlighted on screen with a green frame. Children then labeled the three features of the remaining seven candidate monsters (current monster highlighted with a green frame), starting with the higher or the lower-level feature (counterbalanced). To keep the task design as similar to the initial maze-exploration game, we presented the hypothesis-space spatially as a hierarchical tree (see Figure 8). During the search phase, children were instructed to ask verbal questions to find the target monster and had to pay one star for every question. During feedback, monsters and feature variants were removed from display once they had been eliminated. The same hierarchical and vertical inferences as in the initial maze-exploration game applied. For example, if a higher-level feature had been ruled out (e.g., the target monster is not blue), the experimenter removed all feature variants and candidate monsters connected to the eliminated feature variant from display (e.g., heart pattern, checked pattern, and both blue monsters, see Figure 8).

Results

Mean Information Gain of Participants' Queries

Figure 3a shows the mean information gain of participants' queries in Experiment 3. A linear regression with age (in months) as predictor and mean information gain as dependent variable showed no significant effect of age on the informativeness of children's questions (online supplementary materials, Table S2).

Children's First Queries and Overall Strategy Optimality

Figure 3b shows the proportion of children targeting each level with their first query in Experiment 3. Overall, 13.0% of children made the optimal first query (three 4-year-old, two 5-year-olds, two 6-year-olds, see Figure 4a). The number of children employing the optimal overall search strategy was also very low (one 5-year-old and one 6-year-old, see Figure 4b).

Cross-Experiment Comparison of Informativeness of Children's Queries With and Without Training Phase

To assess to what extent differences in the amount of training may account for children's superior performance in the maze-exploration game relative to the 20-questions game, we compared the mean information gain children achieved in the 20-questions game in Experiment 3 (i.e., question asking with familiarization), in Experiment 1 (i.e., question asking without familiarization) and the maze-exploration game in Experiment 1. We built a mixed linear regression model with task (question asking with familiarization vs. question asking without familiarization vs. maze-exploration game) and age (in months)[5] as predictors and mean information gain as dependent variable. This showed no performance differences between the two 20-questions games of Experiments 1 and 3. However, children achieved significantly higher mean expected information gain in the maze-exploration game than in the 20-questions game with familiarization (Experiment 3), suggesting that adding a familiarization phase to the 20-questions game did not significantly improve their performance on this task and that children's better performance on the maze-exploration game cannot be accounted for by the additional training they received (online supplementary materials, Table S8).

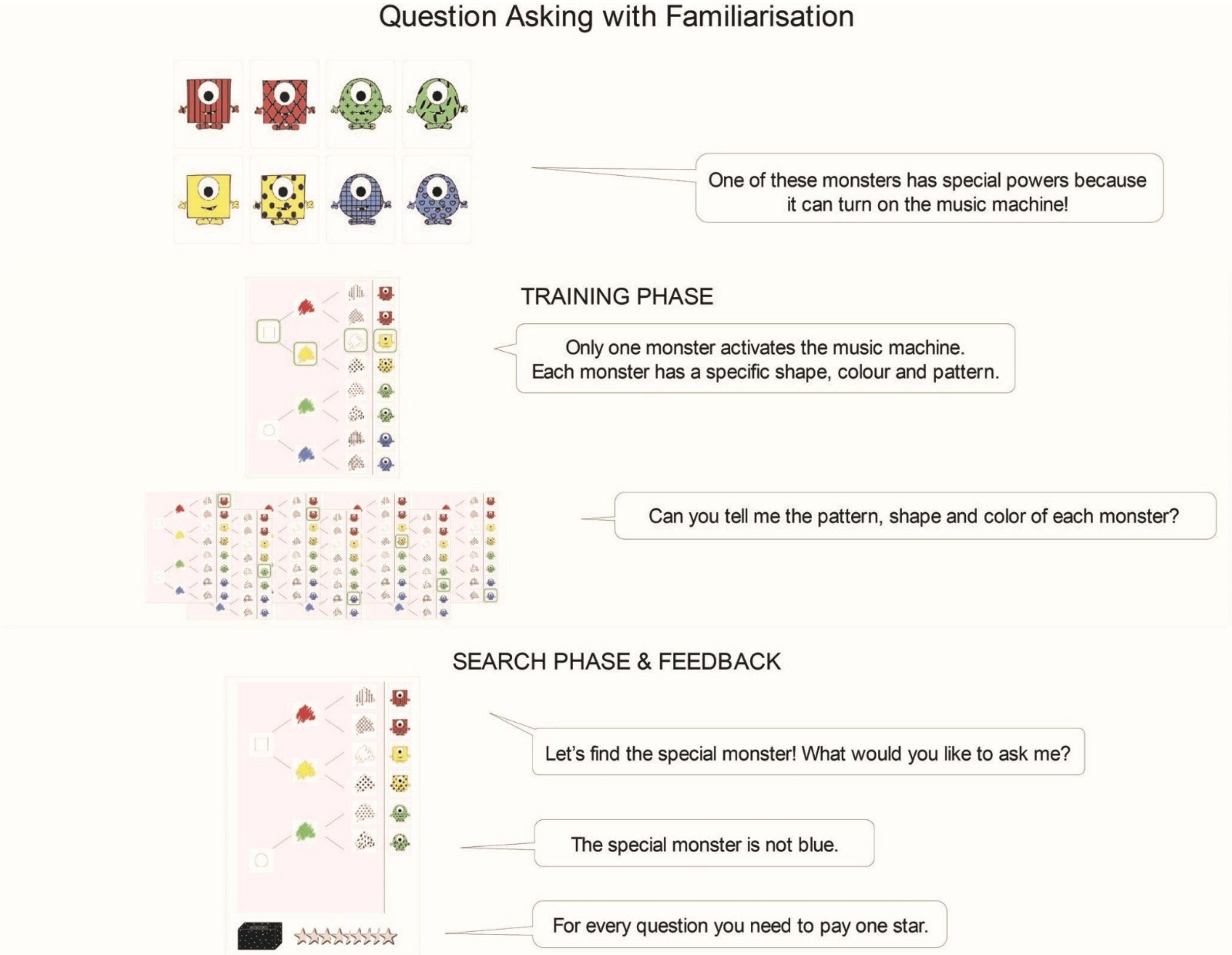
General Discussion

Across several studies, we investigated whether laying out the structure of a search task *spatially* and reducing verbal demands associated with question generation improves the efficiency of 4- to 7-year-old children's information search strategies. To do so, we contrasted children's performance on a novel spatial search task (maze-exploration game) with their information-search behavior in different versions of a structurally and computationally analogous hierarchical 20-questions game. In Experiment 1, we found that, as predicted, children overall made more informative queries in the maze-exploration game than in the traditional 20-questions game. The results of Experiment 1 also revealed a

[4] Seven-year-olds tested in Experiments 1 and 2a were excluded from the analysis, as 7-year-olds were not included in the sample of Experiment 2b.

[5] See footnote 4.

Figure 8
Procedure of Experiment 3



Note. Children played one 20-questions game and completed a familiarization phase prior to the search phase. They had to ask yes-no questions to find the monster, and paid one sticker for every question asked. Sessions were conducted via an online videoconferencing platform. See the online article for the color version of this figure.

developmental increase in the informativeness of children's questions in the 20-questions game. However, there were no age-related performance differences in the maze-exploration game, with younger children's performance on par with that of older ones. In Experiment 2a, children achieved the same performance on the feature-selection version of the 20-questions game as on the maze-exploration game, and older children made more informative queries than younger ones. Experiment 2b showed that providing children with a visual conceptual aid, breaking down the hypothesis space into its categories, was not sufficient to improve their search performance relative to the original 20-questions game. Experiment 3 ruled out the alternative hypothesis that children performed better in the maze-exploration game because they completed an extensive training phase in the maze, but not the 20-questions game. When considered together, the results of our studies suggest that merely providing children with a visual representation of the hypothesis space and/or familiarizing them with the hypothesis-space structure prior to the search phase was not sufficient to improve performance. On the other hand, children did make more informative queries when they were able to make them by pointing rather than through verbal questions.

Several conclusions can be drawn from these results: First, the finding that 4-year-olds performed as well as much older children in our spatial search task and the feature-selection game indicates that even young children possess more effective and sophisticated information-search strategies than previously assumed—at least in tasks that do not require them to generate verbal questions. This result therefore adds to a growing body of research pointing toward the early emergence of effective search and exploration strategies (e.g., Bonawitz et al., 2012; Cook et al., 2011; Domberg et al., 2020; Kushnir & Gopnik, 2005;

Ruggeri et al., 2017; Ruggeri et al., 2019; E. Schulz et al., 2019; L. E. Schulz & Bonawitz, 2007) and supports the view that children are skilled active learners from a very young age (e.g., Gopnik et al., 1997; Gopnik & Wellman, 2012).

Second, our results strongly suggest that younger children's difficulties in the 20-questions game are directly attributable to the verbal requirements related to question generation (e.g., Ruggeri et al., 2017). More specifically, they point toward a crucial role of syntax development in the emergence of efficient question-asking, as indicated by the existence of a transitional phase (Experiment 2a) between pointing and asking complete verbal questions, where children provided feature labels without embedding them into a full question. It may be that during this transitional phase, children already possess the cognitive skills to determine which features they should target, but still fail to translate this understanding into a syntactically correct question that complies with the yes-no format imposed by the 20-questions game.

Our finding that improvements in language skills are closely related to the informativeness of children's questions also shows that the wealth of developmental studies employing the traditional 20-questions game likely paints an incomplete picture of the effectiveness of children's information search strategies—especially in young children, whose verbal skills are still developing. Our findings therefore highlight the importance of investigating children's information search using a broader array of paradigms and tasks that are better tailored to the cognitive abilities and constraints of the targeted age groups, and which are therefore more suitable to capture the early emergence of efficient search.

In our question-asking tasks in Experiment 2b, providing children with a visual and conceptual aid (i.e., displaying the spatial organization of the hypothesis space), thereby making it easier for children to understand and thus exploit its structure, was not sufficient to improve search efficiency—even when children were familiarized with the structure of the display (as in Experiment 3). This appears at odds with previous work showing that scaffolding children's categorization can improve the informativeness of their questions (Ruggeri & Feufel, 2015; Ruggeri et al., 2021; Ruggeri et al., 2019). A possible explanation for this discrepancy is that our visual displays of the hypothesis space (on feature cards or as a hierarchical tree) may have been too complicated for children to parse on their own. Compared with other studies, where children were actively encouraged to attend to the different categories and features by the experimenter, they may also not have been engaging or interactive enough to sufficiently attract children's attention. More generally, our results therefore also support the claim that domain-general cognitive processes such as metacognition and uncertainty monitoring (e.g., knowing what information to attend to, being able to monitor one's own knowledge states or knowing when to stop a search, see Lyons & Ghetti, 2011, 2013; Mata et al., 2011; Ruggeri et al., 2016; Salles et al., 2016), as well as more general executive functions (e.g., directing attention toward relevant information, inhibiting unwanted responses, see Best & Miller, 2010; Betsch et al., 2018; Legare et al., 2013) also contribute to the development of efficient information search and question asking (e.g., Jones et al., 2020; Ronfard et al., 2018).

As is often the case with stimuli and scenarios used in developmental psychology research, our tasks have limited ecological validity, as they are more structured and overly simplified than comparable problems children may encounter in real life. In our spatial search task, we artificially constrained the structure of the maze to make it comparable with the 20-questions game (i.e., only one path leading through the maze, only one open passage per level). In addition, children received extensive assistance with the appropriate if-then inferences upon receiving feedback (e.g., learning whether a passage was open or closed in the maze, or learning that the target monster is not blue). As our pilot revealed, without this assistance children often failed to make even the easiest inferences (e.g., if the right passage on the higher level was open, the left passage on the higher level must be closed). In real life, children would be required to make these inferences themselves, which would make these tasks considerably more complicated, especially for younger age groups.

Also, in real-world situations children would rarely be given cue cards to assist them with their queries, as they were in Experiments 2a, 2b, and 3. Finally, the quantitative framework we used to assess question informativeness can be applied exclusively to environments where only yes-no questions can be asked, and where the hypothesis space is both constrained and predetermined. This setup is not particularly representative of naturalistic conversations, where much richer questions can be asked and where the hypotheses actually considered, as well as their likelihoods, are shaped by the question asker's prior experience and beliefs (Jones et al., 2020). Despite these limitations, the 20-questions game is a classic example of sequential, binary information search, a problem that is actually a very general one encountered throughout the life span, for instance during real-world decision making, medical diagnoses (e.g., Berretty et al., 1997; Green & Mehr, 1997; Hamilton, 2003) or when making causal inferences (see Berretty et al., 1997; Berretty et al., 1999; Martignon et al., 2008). In that sense, the 20-questions game provides a good trade-off between experimental tractability and real-world generalizability.

Despite the limited ecological validity of our maze-exploration game in its current form, we believe it to be a very promising paradigm. Indeed, similar maze paradigms have previously been used successfully with even younger children and primates to study their ability to plan ahead (Völter & Call, 2014). Our addition of an information search component opens up a range of possibilities for the development of nonverbal active search tasks. Using a combination of behavioral analyses and computational modeling, future work could for instance investigate how children explore more complex mazes, and search through maze-like configurations in real-life search paradigms. Building on previous work (Völter & Call, 2014), our paradigm may also be used to study how children and adults search for information to plan ahead, for example to examine whether effective information search contributes to successful planning. In addition, this paradigm provides possibilities for studying toddlers' ecological learning and adaptiveness (Ruggeri et al., 2015, 2017; Ruggeri et al., 2019), dynamic learning over trials, as well as exploration versus exploitation strategies in spatial (Meder et al., 2020; E. Schulz et al., 2019) and nonspatial domains (Wu et al., 2020). Finally, future work may also attempt to further reduce verbal demands in spatial search tasks, for instance using eye tracking or EEG measures, to explore active information search in infants. Employing experimental paradigms that minimize additional demands, such as verbal skills, can provide a much more nuanced understanding of young children's information-search abilities. This approach is therefore crucial to enrich our understanding of the developmental trajectory of children's active learning strategies and to paint a more comprehensive picture of young children's cognitive development.

References

Berretty, P. M., Todd, P. M., & Blythe, P. W. (1997). Categorization by elimination: A fast and frugal approach to categorization. In M. G. Shafto, & P. Langley (Eds.), *Proceedings of the Nineteenth Annual Conference of the Cognitive Science Society* (pp. 43–48). Lawrence Erlbaum Associates.

Berretty, P. M., Todd, P. M., & Martignon, L. (1999). Categorization by elimination: Using few cues to choose. In G. Gigerenzer, P. M. Todd, & ABC Research Group (Eds.), *Simple heuristics that make us smart* (pp. 235–254). Oxford University Press.

Best, J. R., & Miller, P. H. (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., Lehmann, A., Jekel, M., Lindow, S., & Glöckner, A. (2018). Children's application of decision strategies in a compensatory environment. *Judgment and Decision Making, 13*(6), 515–528.

Bonawitz, E. B., van Schijndel, T. J., 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

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

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

Denney, D. R., Denney, N. W., & Ziobrowski, M. J. (1973). Alterations in the information-processing strategies of young children following observation of adult models. *Developmental Psychology, 8*(2), 202–208. https://doi.org/10.1037/h0034144

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

Domberg, A., Koksuba, K., Rothe, A., & Ruggeri, A. (2020). Goal-adaptiveness in children's cue-based information search. In S. Denison, M. Mack, Y. Xu, & A. B.C (Eds.), *Proceedings of the 42nd annual conference of the cognitive science society* (pp. 1437–1443). https://cogsci.mindmodeling.org/2020/papers/0298/0298.pdf

Eimas, P. D. (1970). Information processing in problem solving as a function of developmental level and stimulus saliency. *Developmental Psychology, 2*(2), 224–229. https://doi.org/10.1037/h0028746

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175–191. https://doi.org/10.3758/BF03193146

Gigerenzer, G., Todd, P. M., & ABC Research Group. (1997). *Simple heuristics that make us smart.* Oxford University Press.

Gopnik, A., Meltzoff, A. N., & Bryant, P. (1997). *Words, thoughts, and theories (Vol. 1)*. MIT Press.

Gopnik, A., & Wellman, H. M. (2012). Reconstructing constructivism: Causal models, Bayesian learning mechanisms, and the theory theory. *Psychological Bulletin, 138*(6), 1085–1108. https://doi.org/10.1037/a0028044

Green, L., & Mehr, D. R. (1997). What alters physicians' decisions to admit to the coronary care unit? *The Journal of Family Practice, 45*(3), 219–226.

Hamilton, G. C., Sanders, A., Strange, G. S., & Trott, A. T. (2003). *Emergency medicine: An approach to clinical problem solving* (2nd ed.). Saunders Co.

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

Jones, A., Pachur, T., Markant, D. B., & Ruggeri, A. (2020). How the hypothesis space is represented: Evidence from young children's active search and predictions in a multiple-cue inference task. *OSF Preprints*. https://doi.org/10.31219/osf.io/wh2pr

Jones, A., Swaboda, N., & Ruggeri, A. (2020). Developmental changes in question-asking. In L. P. Butler, S. Ronfard, & K. H. Corriveau (Eds.), *The questioning child: Insights from psychology and education* (pp. 118–143). Cambridge University Press. https://doi.org/10.1017/9781108553803.007

Kachergis, G., Rhodes, M., & Gureckis, T. (2017). Desirable difficulties during the development of active inquiry skills. *Cognition, 166*, 407–417. https://doi.org/10.1016/j.cognition.2017.05.021

Kushnir, T., & Gopnik, A. (2005). Young children infer causal strength from probabilities and interventions. *Psychological Science, 16*(9), 678–683. https://doi.org/10.1111/j.1467-9280.2005.01595.x

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

Lyons, K. E., & Ghetti, S. (2011). The development of uncertainty monitoring in early childhood. *Child Development, 82*(6), 1778–1787. https://doi.org/10.1111/j.1467-8624.2011.01649.x

Lyons, K. E., & Ghetti, S. (2013). I don't want to pick! Introspection on uncertainty supports early strategic behavior. *Child Development, 84*(2), 726–736. https://doi.org/10.1111/cdev.12004

Martignon, L., Katsikopoulos, K. V., & Woike, J. K. (2008). Categorization with limited resources: A family of simple heuristics. *Journal of Mathematical Psychology, 52*(6), 352–361. https://doi.org/10.1016/j.jmp.2008.04.003

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

Meder, B., Nelson, J. D., Jones, M., & Ruggeri, A. (2019). Stepwise versus globally optimal search in children and adults. *Cognition, 191*, 103965. https://doi.org/10.1016/j.cognition.2019.05.002

Meder, B., Wu, C. M., Schulz, E., & Ruggeri, A. (2020). Development of directed and random exploration in children. *PsyArXiv*. https://doi.org/10.31234/osf.io/7zhxv

Mosher, F., & Hornsby, J. (1966). On asking questions. In J. Bruner, R. Oliver, & P. Greenfield (Eds.), *Studies in cognitive growth* (pp. 86–102). Wiley.

Nelson, J. D., Divjak, B., Gudmundsdottir, G., Martignon, L. F., & Meder, B. (2014). Children's sequential information search is sensitive to environmental probabilities. *Cognition, 130*(1), 74–80. https://doi.org/10.1016/j.cognition.2013.09.007

Nelson, J. D., Meder, B., & Jones, M. (2018). Towards a theory of heuristic and optimal planning for sequential information search. *PsyArXiv*. https://doi.org/10.31234/osf.io/bxdf4

Ronfard, S., Zambrana, I. M., Hermansen, T. K., & Kelemen, D. (2018). Question-asking in childhood: A review of the literature and a framework for understanding its development. *Developmental Review, 49*, 101–120. https://doi.org/10.1016/j.dr.2018.05.002

Ruggeri, A., Lombrozo, T., Griffiths, T. L., & Xu, F. (2015). Children search for information as efficiently as adults, but seek additional confirmatory evidence. *Proceedings of the 37th Annual Meeting of the Cognitive Science Society* (pp. 2039–2044).

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*, 104013. https://doi.org/10.1016/j.cognition.2019.104013

Ruggeri, A., Walker, C. M., Lombrozo, T., & Gopnik, A. (2021). How to help young children ask better questions? *Frontiers in Psychology, 11*, 586819. https://doi.org/10.3389/fpsyg.2020.586819

Ruggeri, A., Xu, F., & Lombrozo, T. (2019). Effects of explanation on children's question asking. *Cognition, 191*, 103966. https://doi.org/10.1016/j.cognition.2019.05.003

Ruggeri, A., & Feufel, M. A. (2015). How basic-level objects facilitate question-asking in a categorization task. *Frontiers in Psychology, 6*, 918. https://doi.org/10.3389/fpsyg.2015.00918

Ruggeri, A., & Lombrozo, T. (2014). Learning by asking: How children ask questions to achieve efficient search. *Proceedings of the 36th Annual Meeting of the Cognitive Science Society* (Vol. 36, pp. 1335–1040).

Salles, A., Ais, J., Semelman, M., Sigman, M., & Calero, C. I. (2016). The metacognitive abilities of children and adults. *Cognitive Development, 40*, 101–110. https://doi.org/10.1016/j.cogdev.2016.08.009

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

Völter, C. J., & Call, J. (2014). Younger apes and human children plan their moves in a maze task. *Cognition, 130*(2), 186–203. https://doi.org/10.1016/j.cognition.2013.10.007

Wu, C. M., Schulz, E., Garvert, M. M., Meder, B., & Schuck, N. W. (2020). Similarities and differences in spatial and non-spatial cognitive maps. *BioRxiv*. https://doi.org/10.1101/2020.01.21.914556