### Title
Reuse and Remixing in Question Asking Across Development




### Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 46(0)



### Authors
Liquin, Emily G
Rhodes, Marjorie
Gureckis, Todd M



### Publication Date
2024

Peer reviewed

# Reuse and Remixing in Question Asking Across Development

**Emily G. Liquin (emily.liquin@nyu.edu)**
**Marjorie Rhodes (marjorie.rhodes@nyu.edu)**
**Todd M. Gureckis (todd.gureckis@nyu.edu)**

Department of Psychology, New York University
New York, NY 10003 USA

## Abstract

Question asking is a key tool for learning, especially in childhood. However, formulating good questions is challenging. In any given situation, many questions are possible but only few are informative. In the present work, we investigate two ways 5- to 10-year-olds and adults simplify the challenge of formulating questions: by reusing previous questions, and by remixing components of previous questions to form new questions. Our experimental results suggest that children and adults reuse and remix questions and adaptively modulate reuse depending on how informative a question will be in a particular situation. This work shows that task-relevant experience asking questions provides fodder for future questions, simplifying the challenge of inquiry and enabling effective learning.

**Keywords:** question asking; development; information search; expected information gain; learning

## Introduction

Starting early in development, children ask questions to gain information (Chouinard, 2007; Frazier et al., 2009; Ronfard et al., 2018). Question asking is a powerful tool for learning because questions can efficiently target information that could be difficult to obtain otherwise. For example, children ask questions to learn about causal mechanisms (Callanan & Oakes, 1992), unobservables (Fitneva et al., 2013), and generic concepts (Gelman et al., 2008).

With the great power of questions comes great complexity. After identifying uncertainty to be resolved, a learner *could* ask an infinite number of questions. For example, imagine a child who wants to know what is inside their carefully-wrapped birthday present. The child could ask their parent: "Is it a toy?", "Is it something I asked for?", "How much did it cost?", "What is two plus two?", and so on. These questions range from useful to irrelevant. Given the large number of possibilities, how do we find questions that efficiently help us resolve uncertainty and learn about the world?

In this paper, we investigate how prior questions provide fodder for future questions. For example, if a child previously asked "What sound does a dog make?", they might think of this question later when encountering a novel animal (e.g., "What sound does a goat make?"). We call this **reuse**: repeating questions with the same meaning and structure, applied to new situations. In addition, the child might ask similar questions that incorporate some (but not all) elements of a prior question's meaning and structure, like "Does a goat make the same sound as a sheep?" We call this **remixing**: using components of previous questions to generate novel questions.

## Question Asking, Search, Reuse, and Remixing

Prior work on children's question asking has shown that children ask a range of sophisticated questions in naturalistic settings and about realistic stimuli (e.g., Callanan & Oakes, 1992; Chouinard, 2007; Greif et al., 2006). In addition, research has asked about children's capacity to formulate informative questions to resolve uncertainty (e.g., Herwig, 1982; Mosher & Hornsby, 1966), often using computational models to precisely characterize optimal question asking (e.g., Nelson et al., 2014; Ruggeri & Lombrozo, 2015; Ruggeri et al., 2016). Broadly, this work has shown that children sometimes struggle to formulate optimally informative questions.

Interestingly, adults and children are better at identifying informative questions from a list of options than they are at generating informative questions themselves (Rothe et al., 2018; Ruggeri et al., 2017). This could suggest that the main bottleneck in asking good questions is not difficulty evaluating the quality of possible questions, but rather difficulty thinking of possible questions to evaluate. This challenge is not unique to question asking. Many of the problems humans solve every day require searching a large space of possibilities to generate candidate options. For example, deciding what to cook for dinner, naming a pet, and generating a hypothesis all require a similar process.

In these other domains, the possibilities people have considered in the past exert some influence on the possibilities they consider in the present. When solving problems, people are likely to think of previous solutions (Bear et al., 2020; Morris et al., 2021; Phillips et al., 2019). When generating hypotheses, people are biased by hypotheses they have considered in the past (Bonawitz et al., 2014; Dasgupta et al., 2018; Zhao et al., 2024). And when imagining creative products, people incorporate elements of known entities (Smith et al., 1993; Ward, 1994, 2008).

In the present research, we investigate whether similar mechanisms underlie question asking. In particular, we investigate reuse—asking a question that has the same meaning and structure as a previous question—and remixing—repurposing components of a previous question to formulate a new question. Because question asking is an open-ended, generative task, it is difficult to study empirically. However, studying reuse and remixing may provide traction on understanding where questions come from, enabling us to explain



In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.* 

why people ask particular questions in particular situations.

Prior work (Liquin & Gureckis, 2022) provided preliminary support for reuse and remixing in adults. After being exposed to one set of questions, adults often asked questions that either (1) had the same meaning/structure as those questions (reuse), or (2) used components of those questions (remixing). However, several questions remain unanswered.

First, it is unknown whether these strategies underlie children's question asking—and if they do, how children's use of these strategies compares to adults'. For example, reuse and remixing might be even more prevalent in children's question asking relative to adults' because exploiting old questions is computationally simpler than searching for new questions. Consistent with this, children's hypothesis testing behavior can be more repetitive than adults' (Bramley & Xu, 2023).

On the other hand, children's questions could be less tied to previous questions than adults'. Children are more exploratory than adults (Gopnik, 2020), both in seeking information (Blanco & Sloutsky, 2021; Liquin & Gopnik, 2022; Nussenbaum et al., 2023; Schulz et al., 2019) and considering possibilities (Gelpi et al., 2020; Hart et al., 2022; Lucas et al., 2014). We might expect this exploratory tendency to extend to question asking. This could suggest that children search broadly through the space of possible questions, avoiding reuse and remixing.

These conflicting possibilities motivate our central question: do children reuse and remix questions, and how does their use of these strategies compare to adults'?

We also consider when reuse and remixing might be used. Prior work on decision making (Morris et al., 2021) points to (at least) two important considerations: (1) an option's previous quality, and (2) an option's current quality. If a question was particularly informative in the past, it might be more likely to come to mind again in the future—leading to higher levels of reuse and remixing for previously informative questions. Nonetheless, if a question is not informative in the *current* situation (e.g., "What sound does it make?" when learning about piece of furniture), it should not be reused—even if it was informative in the past. Liquin and Gureckis (2022) found only mixed evidence that adults' reuse was modulated according to a question's current informativeness, and they did not investigate past informativeness. Thus, these predictions remain largely untested, especially in children.

## The Present Research

In the present research, we addressed two questions. How do reuse and remixing compare across development? And when do children and adults reuse and remix questions?

To answer these questions, we developed a new question asking task (see Fig. 1), which builds upon established methods for studying question asking in children and adults (e.g., Rothe et al., 2018; Ruggeri et al., 2016). Participants' goal is to identify the features of three hidden monsters (one blue, one red, one purple). Monsters vary in their head shape (square or circle) and number of legs (one, two, or three). After partial information is revealed (i.e., some heads and/or legs are uncovered), participants are prompted to ask the best question they can to help identify the hidden monsters. By changing which partial information is revealed, we can vary which questions are most informative, using optimal models of question informativeness from prior work (e.g., Liquin & Gureckis, 2022; Ruggeri et al., 2016). In a pilot study, we verified that this task elicited questions that were both informative and semantically diverse from 5- to 10-year-olds and adults. In addition, these questions could be formalized as programs in a domain-specific language (see Table 1), following the computational approach proposed by Rothe et al. (2017). Formalizing questions as programs allows a precise definition of reuse and remixing (see Methods).

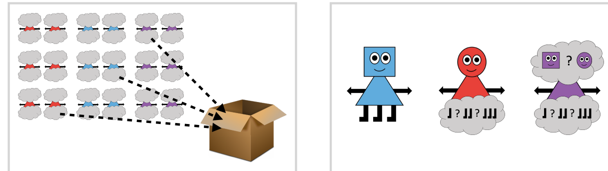


Figure 1: Example trial of question asking task.

In the present research, we experimentally manipulated whether children and adults were exposed to a particular "target question," and then tested whether they later asked questions with the same meaning/structure (reuse) and asked similar, but novel, questions that draw upon components of the target question (remixing). If children and adults reuse and remix questions, we would expect question exposure to affect later question asking. If children differ from adults in their use of reuse and remixing, we would expect an interaction between age and question exposure. We also manipulated the previous and current informativeness of each target question, expecting higher levels of reuse and remixing for target questions that were previously informative, and higher levels of reuse for target questions that were currently more informative. We focused on 5- to 10-year-olds because this is a period of significant developmental change in the ability to ask informative questions (Jones et al., 2020).

The study was preregistered. Preregistration, data, and analysis scripts are available at https://osf.io/4khr9/. Experiment code and online task are available here.

Table 1: Examples of questions and program representations.

| Question | Program Representation |
|---|---|
| How much legs does the purple monster have? | `(legs Purple)` |
| Does any of them have two legs? | `(any (map (λ x0 (== (legs x0) 2)) (set Blue Red Purple)))` |
| Does the purple monster have a square head and two legs? | `(and (== (legs Purple) 2) (== (shape Purple) Square))` |

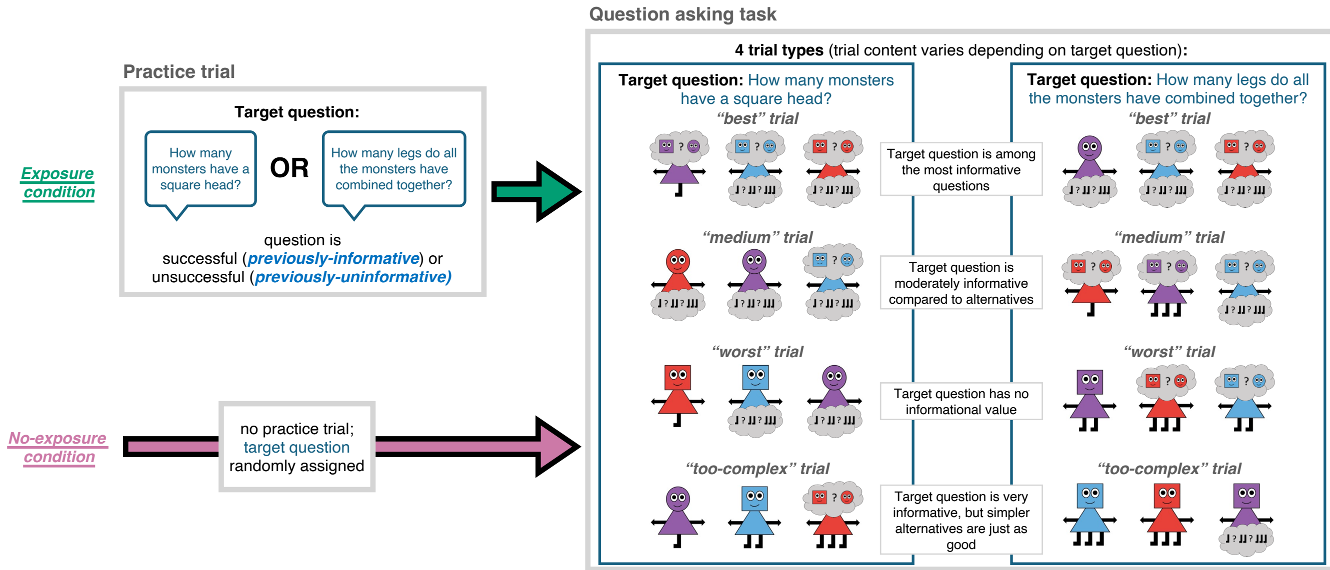


Figure 2: Overview of Procedure.

## Methods

### Participants

Participants were 219 children recruited and tested online through a platform for remote research (66 5- to 6-year-olds, 78 7- to 8-year-olds, and 75 9- to 10-year-olds; 107 boys and 112 girls) and 179 adults recruited and tested online through Prolific (ages 19 to 67; 101 men, 76 women, 2 non-binary/genderqueer). Child participants were 58% White, 17% Asian, 11% multiracial, 5% Black, and 11% Hispanic of any race (2% not specified). Adult participants were 77% White, 9% Black, 4% Asian, 2% multiracial, 1% American Indian/Alaska Native, 8% Hispanic of any race (1% not specified). We excluded an additional 29 adults and 12 children for repeat participation, issues with data saving or video recording, or failure to complete attention/sound checks.

Participants were randomly assigned to an exposure condition (no-exposure, exposure), a previous quality condition (previously-informative, previously-uninformative), and a target question condition ("How many monsters have a square head?" or "How many legs do all the monsters have combined together?"). The exposure condition (which was our main condition of interest) had about twice as many participants ($n = 271$) as the no-exposure condition ($n = 127$). Within each exposure condition, an approximately equal number of participants was assigned to each combination of quality condition and target question condition.

Each participant completed four question asking trials. As preregistered, we excluded all trials that were skipped or had poor audio quality (2% of trials), responses that were non-questions, did not follow the rules, or were off-task (11% of trials), ambiguous questions (6% of trials), and questions that could not be translated into the domain-specific language described below ($<$ 1% of trials). This left a sample of 642 questions from 196 children and 633 questions from 176 adults, which exceeded our preregistered target sample. We excluded more trials from children than adults. In both age groups, exclusion rates were lower on trials that involved low levels of uncertainty ("too-complex trials", see Fig. 2), suggesting that exclusions at least partly reflect task difficulty.

### Procedure

The web version of the task can be viewed `here`. The study was self-guided, consisting of pre-recorded audio narrations and animated videos. Children spoke their responses aloud, and video/audio of the child participating was recorded by the experiment interface. Adults provided typed responses.

First, participants were familiarized with the monsters and relevant features (color, head shape, and number of legs). Next, participants were introduced to the question asking task, in which they would ask a question to identify which monsters were hidden in a box. Questions were subject to one rule: the answer to each question had to be one word (for example, "yes," "square," "two," or "purple"). This prevented questions like, "What do all of the monsters look like?"

Next (see Fig. 2), participants in the exposure condition (but not the no-exposure condition) completed a "practice trial." In this practice trial, a confederate asked a question (the **target question**: "How many monsters have a square head?" or "How many legs do all the monsters have combined together?", depending on target question condition). Participants were prompted to answer the question by choosing from several possible answers. If participants clicked an incorrect answer, they were corrected. Then, the confederate used the answer to guess the hidden monsters.

For each target question, we manipulated whether it was informative or uninformative during the practice trial. In the **previously-informative condition**, it was highly informative based on the partial information provided, and the confederate made a correct guess about the monsters. In contrast, in the **previously-uninformative condition** the question was less informative and the confederate made an incorrect guess. We manipulated both informativeness and guess accuracy because prior research suggests that children sometimes attend to success over informativeness (Török et al., 2023).

Finally, participants completed four question asking trials (see Fig 2). The target question had varying levels of informativeness across the four trials, using optimal models of question informativeness (e.g., Liquin & Gureckis, 2022; Ruggeri et al., 2016). As a result, the question asking trials differed between the two target question conditions. Drawing from all questions asked in a pilot study, the target question was among the most informative questions for one trial **("best" trial)**, a moderately informative question for one trial **("medium" trial)**, and an uninformative question for one trial **("worst" trial)**. For another trial, the question was maximally informative, but more complex than needed **("too-complex" trial)**. The order of the trials was randomized. After each trial, the hidden monsters were revealed.

Though participants in the no-exposure condition did not complete the practice trial and thus were not exposed to a question, the trials they saw in the question asking task were matched to a particular target question. Thus, the "target question" for each participant in the no-exposure condition was defined as the question that matched the trials they saw.

This provides an ideal control condition: if participants ask the target question (or similar questions) in the question asking task for any reason having to do with the structure of the trials, then we would expect participants in *both* conditions to ask these questions at similar rates. However, if reuse and remixing occur, then participants in the exposure condition should ask the target question (or similar questions) more often than those in the no-exposure condition. Thus, to establish reuse and remixing, we compare the no-exposure and exposure conditions. To investigate age differences in reuse and remixing, we test the interaction between exposure condition and age group, indicating whether the size of the condition difference varies across ages.

## Domain-Specific Language

Children's and adults' questions were modeled as programs in a "domain-specific language" adapted from Rothe et al. (2017). This facilitates quantification of reuse and remixing (see below). The domain-specific language, akin to a programming language, represents the semantic meaning of each question as a program, made up of primitive functions and operations. Thus, we refer to this as a question's "program representation." See Table 1 for examples, and see Rothe et al. (2017) for further details. Notably, many questions could be represented several different ways in the domain-specific language. For all questions, we ensured that program representations were consistent across multiple instances of the same question.

## Defining Reuse and Remixing

We defined *reuse* as asking a question that matches the target question in meaning and structure, according to the identity and configuration of functions in the question's program representation. For example, `(== (legs Purple) 2)` and `(== (legs Blue) 3)` use the same functions in the same configuration. Thus, in the "How many monsters have a square head?" target question condition, both this exact question and "How many monsters have a circle head?" matched the target question. In the "How many legs do all the monsters have combined together?" target question condition, only this exact question matched the target question (because the question has no free parameters).

Table 2: Measures of similarity used to quantify remixing.

| Measure | Description |
|---|---|
| Tree edit distance | Number of edits required to get from the target question's program representation to the asked question's program representation (Zhang & Shasha, 1989) |
| Text similarity | Cosine similarity between pretrained Sentence-BERT embedddings (Reimers & Gurevych, 2019) of target question and asked question (standardized so that each program representation was associated with a single unique natural language question) |
| Shared functions | Proportion of functions in the asked question's program representation (e.g., `===`, `shape`) that are also in the target question's program representation |
| Shared arguments | Proportion of arguments in the asked question's program representation (e.g., `Blue`, `Square`) that are also in the target question's program representation |

For all questions that did not match the target question, we also quantified *remixing*. We operationalized remixing as similarity between questions: a question that uses components of the target question would be more similar to the target question, compared to a question that does not. We used four different metrics to quantify similarity, summarized in Table 2. Tree edit distance was preregistered, while the other metrics were exploratory.

# Results

## Statistical Approach

All analyses use mixed-effects regression. We preregistered maximal random effects structure, but maximal models rarely converged. To maintain consistency across analyses, we fit all models with random intercepts only (for participant, target question, and trial type, unless otherwise noted). We estimate statistical significance using likelihood ratio tests. We report contrasts or odd ratios based on estimated marginal means.

We preregistered multi-step analyses of age, investigating both age-related effects within childhood and differences between children and adults. For brevity, and because we did not find evidence for developmental change in reuse/remixing (between children and adults, or within childhood), we report only comparisons between children and adults.

## Do Children and Adults Reuse Questions?

First, we asked whether there was evidence for reuse. We found greater use of target-matching questions in the exposure condition compared to the no-exposure condition (see Fig. 3), with no evidence that this reuse effect varied by age. Specifically, we fit a logistic mixed-effects regression model predicting target question use, with exposure condition (no-exposure, exposure) and age group (children, adults) as fixed effects. This model only included data from the best, medium, and too-complex trials[1]. There was evidence for an effect of exposure condition, $\chi^2(1) = 59.29$, $p < .001$, $OR = 10.29$, 95% CI $[4.86, 21.78]$, as well as an effect of age group, $\chi^2(1) = 18.88$, $p < .001$, $OR = 2.70$, 95% CI $[1.68, 4.34]$. However, adding an interaction between exposure condition and age group did not improve the model fit, $\chi^2(1) = 0.62$, $p = 0.43$. Thus, there was an overall "reuse effect," as evidenced by a difference between the no-exposure and exposure conditions, and adults asked the target question more frequently than children across conditions. However, the size of the condition difference did not vary by age.

## Do Children and Adults Remix Questions?

Next, we asked whether there was evidence for remixing. For some similarity metrics, participants' novel questions resembled the target question more in the exposure condition than the no-exposure condition (see Fig. 4). There was no evidence that these remixing effects varied by age. Specifically, we fit a series of mixed-effects regression models predicting similarity to the target question, with exposure condition (no-exposure, exposure) and age group as fixed effects. These models only included data from questions that *did not* match the target question. There was evidence for an effect of exposure condition for two similarity measures: text similarity, $\chi^2(1) = 8.25$, $p = .004$, $\Delta EMM = 0.02$, 95% CI $[0.01, 0.03]$, and shared functions, $\chi^2(1) = 16.02$, $p < .001$, $\Delta EMM = 0.08$, 95% CI $[0.04, 0.11]$. For both measures, there was also an effect of age group, $ps < .001$: adults' questions were more similar to the target question than children's questions. However, an interaction between exposure condition and age group did not improve model fit for either text similarity, $\chi^2(1) = 0.59$, $p = 0.44$, or shared functions, $\chi^2(1) = 0.22$, $p = 0.64$. Thus, though adults' questions were more similar to the target question overall, the size of the remixing effect did not vary between children and adults.

There was no evidence for an effect of exposure condition (at the preregistered $p < .01$ level) on tree edit distance ($p = .02$) or shared arguments ($p = .14$).

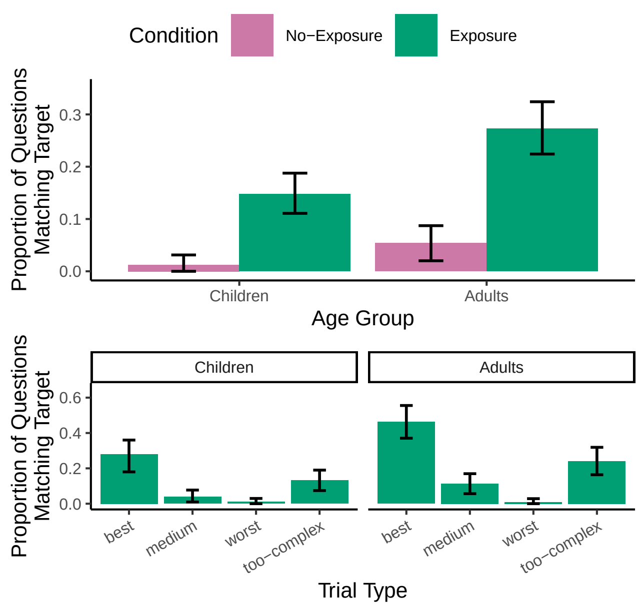


Figure 3: Top: Frequency of target-matching questions (with 95% CIs), across exposure conditions and age groups (only for the best, medium, and too-complex trial types). Bottom: Frequency of target-matching questions (with 95% CIs), across trial types and age groups (exposure condition only). In both plots, y-axes are truncated for ease of visualization.

[1]We strongly expected no use of target-matching questions on the worst trial, where the question was uninformative. Indeed, there were only two instances of target-matching questions on this trial.

## When Do Children and Adults Reuse and Remix?

Next, we investigated *when* participants reused and remixed, as a function of the target question's previous and current informativeness.

We found no evidence that reuse or remixing were affected by the target question's previous informativeness. For reuse, we fit a regression model predicting target question use, with previous quality condition (previously-uninformative, previously-informative) and age group as fixed effects. This model only included data from the exposure condition. There was no evidence for an effect of previous quality condition, $\chi^2(1) = 0.76$, $p = .38$, $OR = 0.82$, 95% CI $[0.53, 1.27]$. There was an effect of age, $\chi^2(1) = 15.27$, $p < .001$, $OR = 2.42$, 95% CI $[1.54, 3.79]$, but there was no evidence that an interaction between previous quality condition and age group improved model fit, $\chi^2(1) = 0.01$, $p = .91$. Thus, though adults asked target-matching questions more than children (which was also true in the no-exposure condition, so does not reflect higher levels of reuse), the likelihood of asking target-matching questions did not depend on the target question's previous quality.

The results for remixing mirrored those for reuse. We conducted this analysis on the two measures of similarity that produced evidence for remixing: text similarity and shared functions. There was no evidence for an effect of previous quality condition on either text similarity, $\chi^2(1) = 0.10$, $p = .75$, $\Delta EMM = 0.002$, 95% CI $[-0.01, 0.02]$, or shared functions, $\chi^2(1) = 0.04$, $p = .84$, $\Delta EMM = -0.004$, 95% CI $[-0.05, 0.04]$. For both measures of similarity, there was an effect of age group, $ps < .001$. However, there was no evidence for an interaction between previous quality condition and age group for either measure of remixing ($ps > .31$). Therefore, though adults' questions were more similar to the target question than children's questions (which was also true in the no-exposure condition, so does not reflect higher lev-

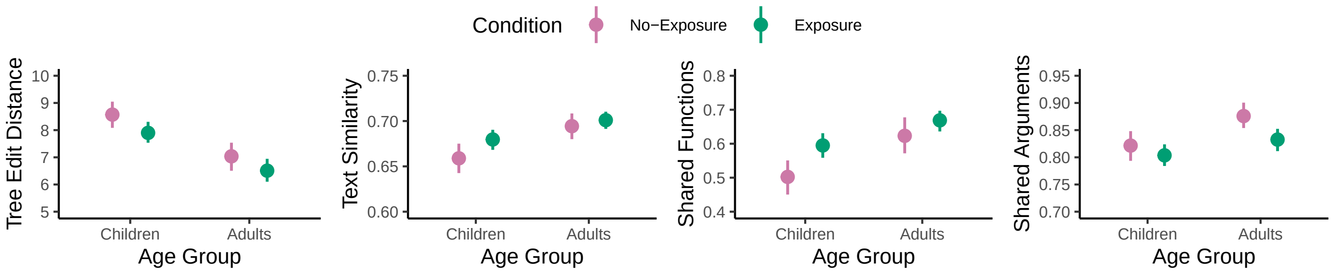


Figure 4: Mean similarity (with 95% CIs) between asked non-target questions and target question, measured by tree edit distance, text similarity, shared functions, and shared arguments, for children and adults in the exposure and no-exposure conditions. For tree edit distance, lower numbers indicate greater similarity. For all other measures, higher numbers indicate greater similarity. Y-axes are truncated from their full ranges for ease of visualization.

els of remixing), there was no evidence that remixing was affected by the target question's previous quality.

Though reuse and remixing were not affected by *prior* informativeness, we found that reuse was more likely when the target question was more informative at the moment of reuse (see Fig. 3). We fit a regression model predicting target question use, with fixed effects for trial type (worst, medium, best, too-complex) and age group (children, adults). This model excluded the random intercept for trial type, and it only included data from the exposure condition. There was a significant effect of both trial type, $\chi^2(3) = 130.72$, $p < .001$, and age group, $\chi^2(1) = 14.38$, $p < .001$. The probability of asking target-matching questions varied across trial types, with higher levels of reuse when the target question would be more informative (e.g., best vs. medium: $OR = 8.87$, 95% CI $[3.81, 20.65]$) but not overly complex (e.g., best vs too-complex: $OR = 2.92$, 95% CI $[1.56, 5.47]$). This effect did not vary across age groups: there was no evidence that adding an interaction between trial type and age group improved model fit, $\chi^2(3) = 0.71$, $p = .87$. Thus, participants reused selectively when it was more informative to do so.

## General Discussion

In this work, we investigated two questions. First, we investigated the developmental trajectory of reuse and remixing in question asking. We found evidence for reuse and remixing in children and adults, but no evidence for developmental differences in the amount of reuse/remixing relative to a control condition. Thus, children do not appear to be more exploratory than adults in their tendency to use prior questions. It is possible that children's question asking is exploratory in other ways, like other kinds of search (Gopnik, 2020).

Second, we asked when children and adults reuse and remix questions. Following prior work on other decision problems (Morris et al., 2021), we investigated two key features: the extent to which a previous question was informative in the past, and the extent to which a previous question would be informative in the present. We found that question reuse was strongly sensitive to current informativeness. This is important because engaging in too much reuse could have informational costs: repeating questions from the past without reflection about one's present situation could lead to asking uninformative questions (e.g., asking "How many monsters have a square head?" when all head shapes are known). Instead, children and adults alike selectively reused questions when it was most informative to do so.

However, we found no evidence that reuse/remixing were more common for previously informative questions compared to previously uninformative questions. Thus, though prior quality shapes what comes to mind in other contexts, this may not extend to question asking. This could be adaptive if the correlation between past and present quality is lower for question asking than for other decision tasks. The dinner recipes that have been good in the past will typically remain good in the present—and events that could change this correlation are infrequent (e.g., becoming a vegetarian). For question asking, in contrast, our knowledge changes with each question we ask. Thus, it may be rare for previously informative questions to remain informative in the present. In future work, it may be fruitful to investigate whether people's sensitivity to past informativeness depends on learned correlations between past and present quality.

This work is limited in that it only investigated two different questions as targets for reuse/remixing. Both questions were fairly complicated and, perhaps relatedly, were less frequently asked by children overall compared to adults. The challenge of reasoning about these questions may have prevented higher levels of reuse and remixing. Therefore, it will be important for future work to extend these findings to a broader set of questions and tasks. In addition, we only provided one exposure trial, which may have weakened any possible effect of previous informativeness.

Despite these limitations, this research provides evidence that children and adults reuse and remix previous questions. They do so in ways that are sensitive to the informational context, reusing most when it is most informative to do so. This work provides new insight into the mechanisms behind question asking and, more broadly, has implications for understanding how humans search through large possibility spaces. Interestingly, question asking does not appear to fully accord with search in other domains. This raises new questions for understanding how humans find good options in a large space of possibilities—whether solving problems, generating hypotheses, or asking questions to learn.

## Acknowledgments

We would like to thank Brenda Chavez-Olguin, Sophia Cordeiro, Alison De Leon, Julia Johnson, Megan Moran, Archisha Murthy, Jasmine Tepper, and Barron Tsai for their assistance with methodology, data collection, and data processing. We are also grateful to members of the Computation and Cognition Lab and the Conceptual Development and Social Cognition Lab at New York University for their valuable feedback on this work. This work was supported by a National Science Foundation SBE Postdoctoral Research Fellowship to E.G.L. under Grant No. 2204021 and National Science Foundation Grant No. 2000617 to M.R.

## References

Bear, A., Bensinger, S., Jara-Ettinger, J., Knobe, J., & Cushman, F. (2020). What comes to mind? *Cognition*, *194*, 104057.

Blanco, N. J., & Sloutsky, V. M. (2021). Systematic exploration and uncertainty dominate young children's choices. *Developmental Science*, *24*(2), e13026.

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.

Bramley, N. R., & Xu, F. (2023). Active inductive inference in children and adults: A constructivist perspective. *Cognition*, *238*, 105471.

Callanan, M. A., & Oakes, L. M. (1992). Preschoolers' questions and parents' explanations: Causal thinking in everyday activity. *Cognitive Development*, *7*(2), 213–233.

Chouinard, M. M. (2007). Children's questions: A mechanism for cognitive development. *Monographs of the Society for Research in Child Development*, *72*(1), 1–129.

Dasgupta, I., Schulz, E., Goodman, N. D., & Gershman, S. J. (2018). Remembrance of inferences past: Amortization in human hypothesis generation. *Cognition*, *178*, 67–81.

Fitneva, S. A., Lam, N. H., & Dunfield, K. A. (2013). The development of children's information gathering: To look or to ask? *Developmental Psychology*, *49*(3), 533–542.

Frazier, B. N., Gelman, S. A., & Wellman, H. M. (2009). Preschoolers' search for explanatory information within adult-child conversation. *Child Development*, *80*(6), 1592–1611.

Gelman, S. A., Goetz, P. J., Sarnecka, B. W., & Flukes, J. (2008). Generic language in parent-child conversations. *Language Learning and Development*, *4*(1), 1–31.

Gelpi, R., Prystawski, B., Lucas, C. G., & Buchsbaum, D. (2020). Incremental hypothesis revision in causal reasoning across development. In S. Denison, M. Mack, Y. Xu, & B. Armstrong (Eds.), *Proceedings of the 42nd annual conference of the cognitive science society* (pp. 974–980). Cognitive Science Society.

Gopnik, A. (2020). Childhood as a solution to explore-exploit tensions. *Philosophical Transactions of the Royal Society B*, *375*, 20190502.

Greif, M. L., Kemler Nelson, D. G., Keil, F. C., & Gutierrez, F. (2006). What do children want to know about animals and artifacts?: Domain-specific requests for information. *Psychological Science*, *17*(6), 455–459.

Hart, Y., Kosoy, E., Liquin, E. G., Leonard, J. A., Mackey, A. P., & Gopnik, A. (2022). The development of creative search strategies. *Cognition*, *225*, 105102.

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.

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.* Cambridge University Press.

Liquin, E. G., & Gopnik, A. (2022). Children are more exploratory and learn more than adults in an approach-avoid task. *Cognition*, *218*, 104940.

Liquin, E. G., & Gureckis, T. M. (2022). Where questions come from: Reusing old questions in new situations. In J. Culbertson, A. Perfors, H. Rabagliati, & V. Ramenzoni (Eds.), *Proceedings of the 44th annual conference of the cognitive science society* (pp. 1160–1167). Cognitive Science Society.

Lucas, C. G., Bridgers, S., Griffiths, T. L., & Gopnik, A. (2014). When children are better (or at least more open-minded) learners than adults: Developmental differences in learning the forms of causal relationships. *Cognition*, *131*(2), 284–299.

Morris, A., Phillips, J., Huang, K., & Cushman, F. (2021). Generating options and choosing between them depend on distinct forms of value representation. *Psychological Science*, *32*(11), 1731-1746.

Mosher, F. A., & Hornsby, J. R. (1966). On asking questions. In J. S. Bruner, R. R. Oliver, P. M. Greenfield, & J. R. Hornsby (Eds.), *Studies in cognitive growth.* 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.

Nussenbaum, K., Martin, R. E., Maulhardt, S., Yang, Y. J., Bizzell-Hatcher, G., Bhatt, N. S., … Hartley, C. A. (2023). Novelty and uncertainty differentially drive exploration across development. *eLife*, *12*, e84260.

Phillips, J., Morris, A., & Cushman, F. (2019). How we know what not to think. *Trends in Cognitive Sciences*, *23*(12), 1026–1040.

Reimers, N., & Gurevych, I. (2019). *Sentence-BERT: Sentence embeddings using siamese BERT-networks.* arXiv.

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.

Rothe, A., Lake, B. M., & Gureckis, T. (2017). Question asking as program generation. In *Advances in neural infor-*

*mation processing systems* (pp. 1046–1055).
Rothe, A., Lake, B. M., & Gureckis, T. M. (2018). Do people ask good questions? *Computational Brain & Behavior*, *1*(1), 69–89.
Ruggeri, A., & Lombrozo, T. (2015). Children adapt their questions to achieve efficient search. *Cognition*, *143*, 203–216.
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.
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.
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.
Smith, S. M., Ward, T. B., & Schumacher, J. S. (1993). Constraining effects of examples in a creative generation task. *Memory & Cognition*, *21*(6), 837–845.
Török, G., Swaboda, N., & Ruggeri, A. (2023). Smart or just lucky? Inferring question-asking competence from strategies' efficiency versus effectiveness. *Developmental Psychology*.
Ward, T. B. (1994). Structured imagination: The role of category structure in exemplar generation. *Cognitive Psychology*, *27*(1), 1–40.
Ward, T. B. (2008). The role of domain knowledge in creative generation. *Learning and Individual Differences*, *18*(4), 363–366.
Zhang, K., & Shasha, D. (1989). Simple fast algorithms for the editing distance between trees and related problems. *SIAM Journal on Computing*, *18*(6), 1245–1262.
Zhao, B., Lucas, C. G., & Bramley, N. R. (2024). A model of conceptual bootstrapping in human cognition. *Nature Human Behaviour*, *8*(1), 125–136.