Effects of explanation on children's question asking

Azzurra Ruggeri[1,2], Fei Xu[3] and Tania Lombrozo[4]

[1] MPRG iSearch, Max Planck Institute for Human Development

[2] School of Education, Technical University Munich

[3] Department of Psychology, University of California, Berkeley

[4] Department of Psychology, Princeton University

(In press, Cognition)

Correspondence concerning this article should be addressed to Azzurra Ruggeri, MPRG iSearch, Max Planck Institute for Human Development, Lenzteallee, 94, 14195 Berlin, Germany. Phone: + 49 152 08808565. E-mail: ruggeri@mpib-berlin.mpg.de

Abstract

The capacity to search for information effectively by asking informative questions is crucial for self-directed learning and develops throughout the preschool years and beyond. We tested the hypothesis that explaining observations in a given domain prepares children to ask more informative questions in that domain, and that it does so by promoting the identification of features that apply to multiple objects, thus supporting more effective questions. Across two experiments, 4- to 7-year-old children (N = 168) were prompted to explain observed evidence or to complete a control task prior to a 20-questions game. We found that prior prompts to explain led to a decrease in the number of questions needed to complete the game, but only for older children (ages 6-7). Moreover, we found that effects of explanation manifested as a shift away from questions that targeted single objects. These findings shed light on the development of question-asking in childhood and on the role of explanation in learning.

Keywords: explanation, question asking, active learning, cognitive development.

Effects of explanation on children's question asking

Asking questions is a key mechanism for learning from knowledgeable others. However, asking *good* questions is no easy task. Prior work has found that young children tend to ask relatively uninformative questions. For example, in the traditional 20-questions game, 4- to 6-year-olds often target single objects or hypotheses rather than categories of objects or hypotheses (e.g., asking "Is it the dog?" versus "Does it have four legs?"; Herwig, 1982; Ruggeri, Walker, Lombrozo, & Gopnik, 2019). Moreover, children's questions are often unnecessary, targeting information redundant with what they already know (Legare, Mills, Souza, Plummer, & Yasskin, 2013; Ruggeri, Lombrozo, Griffiths, & Xu, 2016). In this paper, we test the hypothesis that prompting children to engage in *explanation* prior to a question-asking task can increase the efficiency of their inquiry by helping them appreciate more abstract features or categories that support informative questions.

Development of question asking

To investigate the quality of children's questions, we follow prior research using the 20-questions task (see Herwig, 1982; Mosher & Hornsby, 1966; Nelson, Divjak, Martignon, Gudmundsdottir, & Meder, 2014; Ruggeri, Lombrozo, Griffiths, & Xu, 2016; Ruggeri & Feufel, 2015). In this task, children must identify a target object (or set of objects) by asking as few yes/no questions as possible. Although the 20-questions game is in some ways artificial, it is a classic example of sequential, binary information search, which captures the structure of problems encountered throughout the lifespan. Indeed, many real-world decision-making, categorization, and causal inference tasks have been modeled with fast and frugal trees that involve sequential, binary branching (see Berretty, Todd, & Martignon, 1999; Martignon, Katsikopoulos, & Woike, 2008). For example, in emergency medicine, resident physicians learn to check for the presence or absence of certain physiological changes to rule out lethal conditions that can be associated with a particular complaint (e.g., Green & Mehr, 1997; Hamilton, Sanders, Strange, & Trott, 2003). Thus, studying children's performance on a 20-questions game is a good compromise between experimental tractability and real-world generalizability.

Previous studies have found that preschoolers can successfully *identify* which of two questions is more effective in the sense that it will, on average, yield a more informative answer (Ruggeri, Sim, & Xu, 2017). However, they are unable to reliably *generate* the most effective questions themselves (see Herwig, 1982; Legare et al. 2013; for related findings with older children and adults, see Ruggeri & Feufel, 2015; Rothe, Lake, & Gureckis, 2018). Indeed, although the majority of 4-year-old children's questions are informative (as opposed to redundant or uninformative; see Legare et al., 2013), until age 7 they mostly ask questions that target a single object or hypothesis – so-called hypothesis-scanning questions, such as "Is it the dog?" (Herwig, 1982; Mosher & Hornsby, 1966; Ruggeri & Feufel, 2015; Ruggeri & Lombrozo, 2015). For example, in a traditional version of the 20-questions game, Herwig (1982) found that about 90% of the questions asked by first graders, and 83% of those asked by second graders, were hypothesis-scanning questions. However, by age 10 children predominantly ask constraint-seeking questions, which target categories or features shared by several different hypotheses (e.g., "Does it have four legs?"), which typically increase the efficiency of search (see Ruggeri & Feufel, 2015; Ruggeri & Lombrozo, 2015).

This developmental change in children's questions has often been explained by an increasing ability to identify and generate abstract features that can be used to group similar objects into categories at different levels (e.g., "animals" or "four-legged creatures" versus "that dog," see Ruggeri & Feufel, 2015). Supporting this idea, Legare et al. (2013) found that the ability to flexibly categorize objects along different dimensions correlates with the informativeness of questions asked by 4- to 6-year-olds in a 20-questions game. However, prior attempts to improve children's question-asking strategies have yielded mixed results, and have not always measured effects on task performance (e.g., Courage, 1989; Denney, 1972; Denney, Denney, & Ziobrowski, 1973; Denney & Turner, 1979; Ruggeri, Walker, Lombrozo, & Gopnik, 2018). In an experimental intervention, Ruggeri, Walker, Lombrozo, and Gopnik (2019) found that "scaffolding" such higher-level features (i.e., telling children which object features could be used to categorize the objects at different hierarchical levels) led 6-year-olds to ask more informative questions. However, the scaffolding intervention did not have the same effect on 4- or 5-year-olds. Courage (1989) trained 4-, 5-, and 7-year-old children on a 20-questions game by providing explicit instructions about how to ask constraint-seeking questions over eight rounds of the 20-questions game. After training, 4- and 5-year-olds indeed asked more constraint-seeking questions. However, this study did not analyze the informativeness of the constraint-seeking questions asked (e.g., how many objects they targeted beyond two), nor measure a general performance improvement (i.e., a reduction in the number of questions needed to reach the solution). In another study, Denney (1972) trained 6-year-olds by providing them with explicit examples of adults asking either hypothesis-scanning or constraint-seeking questions while playing a short 20-questions game. However, 6-year-olds' rate of hypothesis-scanning versus constraint-seeking questions was unaffected at posttest. These findings suggest that interventions that reliably boost young children's performance would not only improve upon the current state of the art, but, perhaps more importantly, would shed light on what it is that supports children's developing ability to ask effective questions in early childhood.

The present study

In this paper, we explore *explanation* as a process that might improve the effectiveness of children's information search, focusing on children's question-asking performance, by preparing them to identify higher-order features that can be used to ask more effective questions. Previous research has shown that the process of seeking, generating, and evaluating explanations fosters learning (e.g., Chi, Bassok, Lewis, Reimann, & Glaser, 1989; Chi, de Leeuw, Chiu, & LaVancher, 1994; Lombrozo, 2006, 2016; Siegler, 2002), including in early childhood (Crowley & Siegler, 2003; Legare & Lombrozo, 2014; Schult & Wellman, 1997; Walker et al., 2014, 2017; Wellman, 2011; Wellman & Liu, 2007).

Most relevant for our purposes, prompting adults to explain has been shown to help them discover patterns of features shared across items (Chin-Parker & Cantelon, 2017; Williams & Lombrozo, 2010; see Lombrozo, 2016, for a review), and these findings extend to young children as well. For example, Walker et al. (2017) presented 5-year-old children with four objects that activated a machine, and four objects that did not. As each object was placed on the machine, half of the children were asked to *explain* why the object did or did not activate the machine, while those in a control condition were instead asked to indicate *whether* the object did or did not activate the machine. The objects had several features, only one of which could be used to perfectly differentiate those items that did (vs. did not) activate the machine. The children who were prompted to explain were significantly more likely than those in the control condition to use this feature as a basis for making subsequent inferences about which new objects would or would not activate the machine, suggesting that explanation helped them recognize the distribution of this feature across items.

By prompting children to consider patterns in features across items, explanation can also help young children "go beyond the obvious" to represent more subtle or abstract features of what they are trying to explain. For example, Walker et al. (2014) found that prompting 3- to 5-year-old children to explain why particular blocks did or did not activate a machine made them more likely to later generalize internal parts or category membership on the basis of causal similarity (whether or not the machine was activated) over perceptual similarity (which is what typically drives inferences in this age group; see also Legare & Lombrozo, 2014). In a study with 5- to 6-year-old children, Walker and Lombrozo (2017) found that prompts to explain key elements of a story helped children appreciate an abstract lesson (e.g., the value of collaboration) contained in a concrete instantiation (e.g., creating a more beautiful painting by sharing different colors of paint). And in a study with 4- to 6-year-old children, Walker, Bonawitz, and Lombrozo (2017) found that prompting older – but not younger – children to explain helped them go beyond two salient, independent explanations for why two plants were sick (low light and low water) to a less familiar explanation (soil type) that explained both cases by appeal to a single cause.

Studies with adults help isolate what it is about explanations that might drive these effects. In a classification task with adults, Williams and Lombrozo (2010) prompted participants to explain the category membership of novel category exemplars that could be classified on the basis of a salient but imperfect rule, or on the basis of a more subtle rule that accounted for all items. Mirroring the findings with children, those who were prompted to explain were significantly more likely than those in a variety of control conditions to discover the subtle, diagnostic feature. However, those who were prompted to explain were also more likely than those who were prompted to describe to characterize the features of the exemplars in abstract terms (e.g., “warm ” versus “yellow” color, and “pointy” versus “triangular” shape). Relatedly, Williams, Lombrozo, and Rehder (2013) found that participants who were prompted to explain were more likely to focus on abstract “themes” shared across items, even when the themes were instantiated in the form of distinct features. For example, “heavy insulation” and “drives on glaciers” were distinct features of individual vehicles, but they were associated with the common theme of being designed for cold-weather. Lombrozo (2016) suggests that when prompted to explain, both children and adults seek a broad and simple pattern that underlies what they are trying to explain, and that in the course of doing so, they’re more likely to go beyond the obvious and represent the target of explanation in different ways, facilitating discovery and often abstraction (see also Edwards, Williams, Gentner, & Lombrozo, 2018; Kon & Lombrozo, in press; Wilkenfeld & Lombrozo, 2015).

In light of this research, we predicted that prompting children to explain observations within a target domain would help them identify appropriate features for asking informative questions and therefore search for information more effectively. To illustrate, consider a child faced with the task of identifying which kinds of foods are causing her to have tummy aches. In trying to explain why pizza, a quesadilla, and a cheese sandwich caused a stomach ache, while a hot dog, a bean burrito, and a peanut butter sandwich did not, she will be driven to identify what it is that differentiates the former set from the latter set, ideally by identifying a single feature that all and only the former possess. In this case, this could lead her to identify *cheese* or *dairy* as relevant features. Moreover, reasoning to this solution might have led her to consider (if ultimately reject) a variety of other features that are common across subsets of these cases: whether the food involves bread, whether it is eaten hot or cold, whether it contains tomatoes, and so on. The process of explaining will thereby affect the way she represents the features within this domain, potentially setting her up to ask more effective questions in the future. When another friend complains of a tummy ache, for example, she will be prepared to ask whether it is caused by *dairy* rather than *pizza*.

Finding that prompts to explain affect the quality of subsequent question-asking would be significant for several reasons. First, given that most efforts to train young children to ask more informative questions have met with only modest success, a robust and domain-general training intervention would itself be a contribution. Second, and more important, such an effect would help connect explanation and subsequent information search as linked steps within a broader process of inquiry. To date, explanation and information search have overwhelmingly been studied in isolation, with some notable exceptions linking explanation to exploration (e.g., Bonawitz, van Schijndel, Friel, & Schulz, 2012; Legare, 2012; Legare & Gelman, 2014). But as children navigate the world, these processes are often intertwined, with explanation seeking leading to information search, and new information affecting subsequent explanations. To our knowledge, this paper is the first to consider the effects of explanation on the efficiency of subsequent information search.

Finally, our studies have the potential to shed light on both the development of children's ability to ask effective questions, and on the role of explanation in learning more generally. With regard to the former, our predicted role for explanation would dovetail with claims that children's ability to ask effective questions depends in part on their ability to identify patterns of features useful to flexibly categorize cases at different levels of abstraction. With regard to the latter, the predicted finding would not only support the idea that in the course of explaining children can recognize patterns and abstract features, but that this influences the effectiveness of their subsequent information search and active learning. Moreover, analyzing the kinds of questions children ask can potentially speak to alternative hypotheses about *how* explanation affects cognition. In particular, we can investigate the extent to which explanation affects question-asking by leading children to ask the most efficient questions available (which are typically at a higher level of abstraction), or instead by nudging them away from what seems to be their default strategy of asking questions at the level of individual objects. We explore these possibilities and test our prediction across two experiments.

Experiment 1

In Experiment 1, we presented children with a 20-questions game in which the goal was to determine which kinds of treats give a monster a tummy ache. This task was preceded by a training phase in which children observed instances of treats with the same features that did or did not cause tummy aches for four other monsters. Children in the

Training-with-explanation condition were prompted to explain (without feedback) why particular treats gave each of these four monsters a tummy ache; those in the Training-without-explanation condition were not. We predicted that children in the Training-with-explanation condition would identify shared features and category membership in the course of explaining, and therefore outperform those in the Training-without-explanation condition in the final 20-questions task. Our task also allowed us to explore how prompts to explain affected children's questions: by leading them towards the highest-level questions, or away from the lowest (object-level) questions.

An alternative hypothesis that we considered is that observing structured evidence would be sufficient for children to extract relevant features and categories, even without a prompt to explain. If this is the case, then the Training-without-explanation might be sufficient to improve the quality of young children's questions. For this reason, we included an additional No-training condition in which children completed the 20-questions task without prior training. Including this condition allows us to tease apart the influence of *exposure* to training data from the effects of *explaining* that data.

Method

Participants. Participants were 69 4- to 6-year-old children (35 female, M = 68.8 months; SD = 7.5 months) recruited at local museums in the Bay Area, California. The children were native English speakers or fluent in English; they were predominantly white and belonged to various social classes. Nine additional children were excluded from analyses due to experimental error (N = 5) or because they did not want to continue the experiment (N = 4). Written informed consent was obtained from participants' parents and the local ethical review board at the University of California, Berkeley approved the study protocol (#2010-03-1013). Children received a small gift for their participation.

Design and procedure. Children were introduced to a game in individual sessions: "On the little planet of Apres, there live some monsters. The monsters love to eat yummy treats from Earth. But sometimes these treats give them a tummy ache! They need your help to figure out which kinds of treats give them tummy aches!" Children were randomly assigned to one of three conditions, Training-with-explanation, Training-without-explanation, or No-training. All children then received the same *test trial*, which we describe first.

Test trial. For the test trial, children were presented with Loma, a monster from planet Apres, and 16 cards representing treats from planet Earth that Loma had tried (see Figure 1). Children were told that some treats give Loma a tummy ache, and their job was to figure out what kind of treats give Loma a tummy ache by asking yes/no questions (e.g., "Do cupcakes give Loma a tummy ache?", "Does the lollipop give Loma a tummy ache?"). For all children, the correct solution was that *cupcakes* give Loma a tummy ache.

This task was designed to be a *hierarchical* version of the 20-questions game (developed by Ruggeri et al. 2016), which differs from traditional versions of the 20-questions game in two ways. First, the solution to the problem is a *category* of objects (i.e., "all cupcakes") rather than an individual object (e.g., "this one cupcake"). Thus, the hypotheses to be considered do not correspond to individual objects, and therefore targeting individual objects will be even less effective than in the traditional version of the 20-questions game. Second, the array of objects that participants received could be classified into a symmetrical nested structure organized at three category levels, each including the same number of objects. We refer to these as the higher level (8 cupcakes and 8 other treats), middle level (among the cupcakes, 4 have sprinkles and 4 have cherries; among the other treats, 4 are sweet and 4 are salty), or lower level (cupcakes come in pairs of different colors; among the sweet treats, 2 are candies, and 2 have chocolate; among the salty treats, 2 are in bags, 2 contain meat). This structure was chosen so that we could analyze children's questions in a more fine-grained way: in addition to investigating whether their questions tended to target single objects versus groups of objects, we could see whether they tended to pick features at different levels (lower, middle, or higher).[1] This structure has also been analyzed computationally (Ruggeri et al. 2016), so we know that the most efficient route to the solution is to begin asking questions that target higher levels (i.e., the most informative initial questions target the higher level, followed by the middle level and then the lower level).

As children asked questions to try to identify the solution, the experimenter modified the visual array of cards to help them keep track of what they had learned. Specifically, after answering each question with "yes" or "no," the experimenter moved the treat(s) targeted by the question from the central array to a location under a card representing a happy face (no tummy ache) or under a card representing a sad face (tummy ache; see Figure 1). For example, if a child asked, "Did the treats with cherries give Loma a tummy ache?", the experimenter would respond "yes" and move the four cupcakes with cherries under the sad face.

[1] Based on Ruggeri et al. (2016), we would not expect differences in performance depending on the hierarchical level of the solution to the problem.

The stimuli were designed to minimize the possibility that children would ask cross-category questions—that is, those that targeted groups of objects belonging to different categories. Following previous research with the 20-questions game (see Ruggeri et al., 2016), questions leading to mixed yes/no feedback (i.e., those targeting a mixed set of cupcakes and other treats) would not be answered directly. Instead, the experimenter would say "Mmm…I am not sure about this one. Can you ask me another question?" Note that only two of such questions were asked in Experiment 1, and seven in Experiment 2.

After each response, children were asked if they thought they knew what kind of treats give Loma a tummy ache, or if they wanted to ask another question. After reaching the solution (or having sorted all treats into the happy/sad piles), children were asked: "Why do you think *these* treats (pointing to the ones in the sad pile) give Loma a tummy ache and not *these* (pointing to the ones in the happy pile)?"

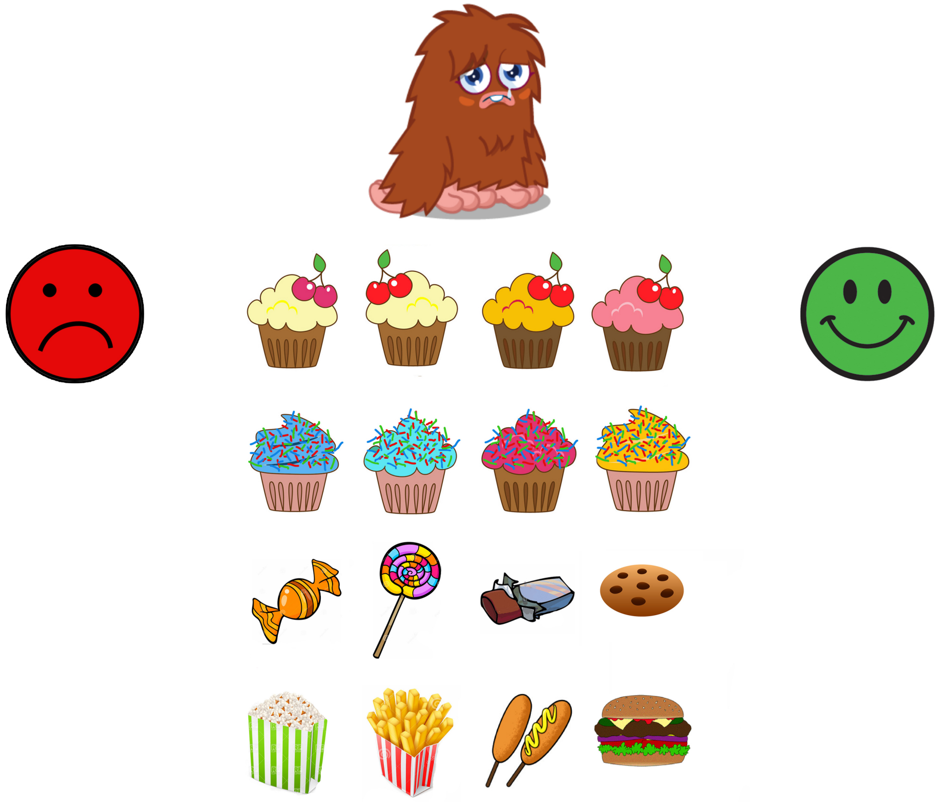

Figure 1. Set up of the test phase common to all conditions.

Training. In the No-training condition, children completed the test trial without a prior task. However, for children in the two training conditions, the test trial was preceded by a training phase in which children observed four monsters, one at a time, presented in random order (see Table 1). For each monster, children were first shown four treats (in random order) that gave that monster a tummy ache, and then four treats (in random order) that did not give the monster a tummy ache. This training was designed to offer children structured data suggesting relevant categorical features at different levels of abstraction. Specifically, one of the monsters (Seria) was presented with data suggesting that the relevant category for tummy aches was cupcakes, another (Polla) was presented with data suggesting that the relevant category was cherries, another (Menna) was presented with data suggesting that the relevant category was sprinkles, and one (Nara) was presented with data suggesting that the relevant category was cherries *or* sprinkles. Note that all relevant categories went beyond a single object, that the categories were at multiple levels of the hierarchical structure (e.g., cherries versus cupcakes), and that the set included “cupcakes,” which was the correct solution to the subsequent test trial.

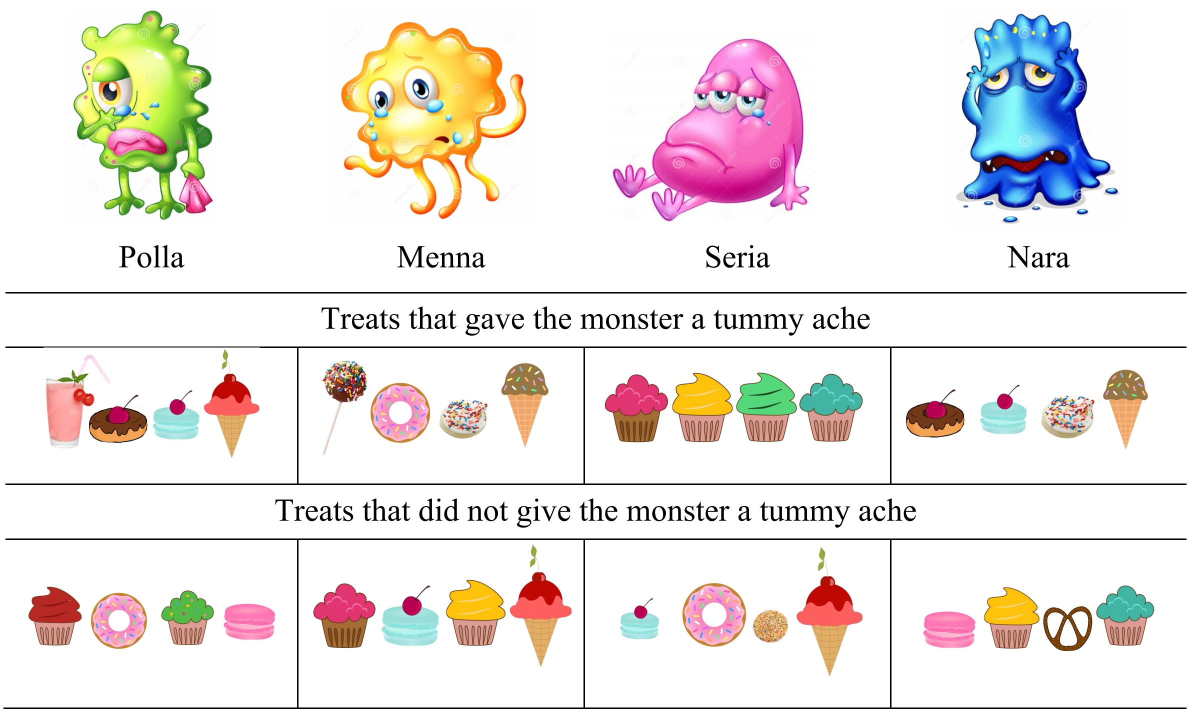

Table 1. Training sets for the Training with explanation and training without explanation conditions.

In the Training-with-explanation condition, children were asked, for each monster, “Why did these treats [pointing to the four treats that gave the monster a tummy ache] give Seria/Menna/Polla/Nara a tummy ache, and not these [pointing to the four treats that did not give the monster a tummy ache]?” The experimenter commented neutrally on all of children’s explanations (“Hmm, that’s an interesting idea”) before moving on. Children who said “I don’t know” were prompted twice again (“What do you think? Why these and not these [pointing to the four treats that gave/did not give the monster a tummy ache]?”). When children did not provide any explanation after prompting, the experimenter said “That’s okay, let’s look at this other monster” and moved on.

In the Training-without-explanation condition, children were asked, for each monster, to “point to the treats that gave Seria/Menna/Polla/Nara a tummy ache” and to “point to the treats that *did not* give Seria/Menna/Polla/Nara a tummy ache.” These prompts were designed to match the explanation condition in drawing attention to the four treats that did and did not cause a tummy ache, and to the *contrast* between these two sets.

Results

The Database containing the data for both Experiments included in this manuscript is archived in a public repository, linked in the Supplementary Materials (Database: Ruggeri, Xu, & Lombrozo, 2019).

Number of questions. We first analyzed the number of questions required to complete the test trial (i.e., to reach the point where every object was classified under the happy face or the sad face) as the dependent variable in a univariate ANOVA with condition as an independent variable. This analysis revealed a main effect of condition, $F(2,68) = 6.69$, $p = .002$, $\eta^2 = .17$. Bonferroni corrected post-hoc analyses confirmed that children in the Training-with-explanation condition needed fewer questions to complete the task ($M = 5.67$; $SD = 3.50$) than did those in the Training-without-explanation condition ($M = 9.96$; $SD = 4.80$, $p = .004$) or the No-training condition ($M = 9.52$; $SD =$

4.32, $p = .010$). There was no difference between the Training-without-explanation and the No-training conditions ($p = 1.00$).

To test for developmental effects, we split children into two age groups at the median age: younger (35 participants, 11 to 12 per condition, $M = 63.1$ months; $SD = 3.4$ month) and older children (34 participants, 10 to 13 per condition, $M = 74.7$ months; $SD = 5.6$ months). A univariate ANOVA with number of questions required to complete the task as the dependent variable and condition and age as independent variable revealed a main effect of condition, $F(2,68) = 7.59$, $p = .001$, $\eta^2 = .19$ (consistent with the previous analysis) and no main effect of age ($p = .445$). However, there was a significant interaction between condition and age, $F(2,68) = 3.43$, $p = .039$, $\eta^2 = .10$. As shown in Figure 2, the explanation prompt had an impact for older children, but not for younger children.

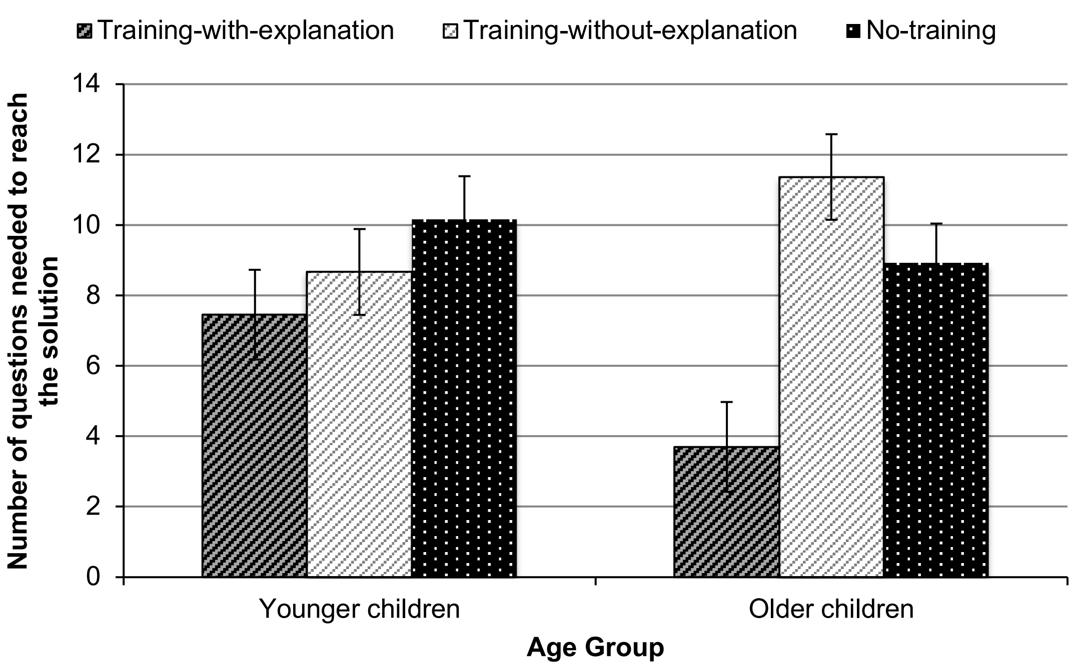

Figure 2. Number of questions children needed to complete the task, displayed by condition (Training-with-explanation, Training-without-explanation and No-training) and

age group (younger and older children, obtained by median split). Bars represent one SEM in each direction.

Univariate ANOVAs within each age group showed that the number of questions needed by younger children did not differ significantly across conditions (p = .317; see Figure 2). However, the number of questions needed by older children did, $F(2,33)$ = 9.30, p < .001, η^2= .39. A Bonferroni corrected post-hoc analysis confirmed that older children needed fewer questions to reach the solution in the Training-with-explanation condition (M = 3.70; SD = 1.89) than in the Training-without-explanation (M = 11.36; SD = 4.41, p < .001) or the No-training conditions (M = 8.92; SD = 4.80, p = .013), which did not differ from each other (p = .448).

Table 2

Mean Percentage (and SDs) of Questions That Targeted The Lower, Middle, Higher Category Levels or Individual Objects in Experiment 1, Displayed By Condition (Training-with-explanation, Training-without-explanation and No-training) and Age Group (Younger and Older Children, Obtained by Median Split).

Age group	Condition	Individual-Object level	Lower level	Middle level	Higher level
Younger children	Training-with-explanation	66 (33)	6 (10)	19 (25)	9 (16)
	Training-without-explanation	71 (21)	6 (11)	16 (21)	6 (12)
	No-Training	79 (23)	6 (10)	7 (18)	7 (14)
Older children	Training-with-explanation	30 (39)	1 (4)	38 (43)	30 (40)
	Training-without-explanation	92 (15)	1 (3)	1 (2)	6 (15)
	No-Training	79 (29)	3 (8)	6 (15)	12 (18)

Kinds of questions asked. To investigate whether and how the explanation training affected the kinds of questions children asked, we classified questions according to the hierarchical level that they targeted: higher (e.g., cupcakes), middle (e.g., cherries), lower (e.g., chocolate), or individual object (e.g., popcorn). This allowed us to calculate, for each child, the percentage of questions asked of each type (see Table 2). We then performed univariate ANOVAs for each question type, with the overall percentage of questions of that type as the dependent variable, and condition and age group as independent variables. These analyses revealed no main effect of condition or interactions on the percentage of questions at the lower (p = .857) or higher (p = .092) levels. However, there were significant effects at both the middle and the individual-object levels.

Analyses at the middle level revealed a main effect of condition, $F(2,68) = 6.09$, $p = .004$, $\eta^2 = .16$. Bonferroni corrected post-hoc analyses showed that children asked a higher percentage of middle-level questions in the Training-with-explanation condition (M = 28%; SD = 35%), as compared to the Training-without-explanation (M = 9%; SD = 17%; p = .017) or the No-training condition (M = 6%; SD = 16%, p = .006). We found no difference between the Training-without-explanation and the No-training conditions (p = 1.00). However, this effect was qualified by an interaction between condition and age, $F(2,68) = 3.11$, $p = .052$, $\eta^2 = .09$. Two univariate ANOVAs separated by age group confirmed that a higher proportion of the older children's questions in the Training-with-explanation condition were at the middle level as compared to the No-training or the Training-without-explanation conditions, $F(2,33) = 6.89$, $p = .003$, $\eta^2 = .31$, whereas for younger children there was no difference across conditions (p = .373; see Table 2).

Analyses at the individual-object level similarly revealed a main effect of condition, $F(2,68) = 9.99$, $p < .001$, $\eta^2 = .24$, qualified by a significant interaction with age group, $F(2,68) = 6.01$, $p = .004$, $\eta^2 = .16$. Bonferroni corrected post-hoc analyses showed that, on average, children asked a lower percentage of individual-object questions in the Training-with-explanation condition (*M* = 49%; *SD* = 39%), as compared to the Training-without-explanation (*M* = 81%; *SD* = 21%; $p < .001$) or the No-training condition (*M* = 79%; *SD* = 26%, $p = .001$). We found no difference between the Training-without-explanation and the No-training conditions ($p = 1.00$). To understand the interaction, we conducted two univariate ANOVAs separated by age group. This confirmed that fewer of the older children's questions in the Training-with-explanation condition were at the individual-object level as compared to the No-training or the Training-without-explanation conditions, $F(2,33) = 13.27$, $p < .001$, $\eta^2 = .46$ (see Table 2). For younger children there were no significant differences across conditions ($p = .470$; see Figure 3).

Figure 3. Percentage of questions children asked that targeted only one treat, displayed by condition (Training-with-explanation, Training-without-explanation and No-training) and age group (younger and older children, obtained by median split). Bars represent one SEM in each direction.

Final explanations. The final explanations provided by children in all conditions were coded as correct when they referred to *all and only* the eight cupcakes (i.e., "cupcakes" or "treats with sprinkles and cherries on top"), and as incorrect otherwise (e.g., "sweet treats"). Explanations were coded by two independent coders, with 100% agreement.

A Chi-square analysis confirmed that the percentage of children providing a correct final explanation did not differ with age (Younger: M = 66%; Older: M = 65%, p = .930) or across the Training-with-explanation (M = 62%), Training-without-explanation (M = 78%) or No-training (M = 56%) conditions (p = .251).

Explanations during training. For children in the Training-with-explanation condition, the four explanations offered during the training phase were coded by two independent coders, with 100% agreement. Answers were coded as "correct" if they identified features that differentiated the four positive and negative examples (e.g., referring to "cherries" for Polla). Children received scores between 0 and 4.

Overall, children provided 1.62 (SD = 1.60) correct explanations. Older children provided more correct explanations (M = 2.30, SD = 1.64) than younger children (M = 1.00, SD = 1.34; $t(19)$ = -2.00, p = .060). Moreover, we found a marginal negative correlation between the number of correct explanations offered during training and the total number of questions needed to reach the solution in the test trial (Pearson's r = -.382, p = .087), suggesting that children who produced more accurate explanations were subsequently more efficient. We also found that children who mentioned "cupcakes" during training needed fewer questions to reach the solution on the test trial: those children who mentioned cupcakes on average needed 4.00 questions (SD = 2.36), whereas those who did not mention cupcakes required an average of 7.18 questions (SD = 3.76), $t(21)$ = -2.29, p = .033).

Finally, there was a significant correlation between the number of correct explanations offered during training and the ability to provide a correct final explanation in the test phase: those children who provided a correct final explanation on average provided 2.15 correct explanations during training (*SD* = 1.52), whereas those who provided an incorrect final explanation on average provided 0.75 correct explanations during training (*SD* = 1.39), $t(21) = 2.12$, $p = .047$). Not surprisingly, those children who mentioned "cupcakes" during training were also more likely to offer a correct final explanation in the test phase (80%) than those who did not (46%), but this difference was not significant ($p = .114$).

Discussion of Experiment 1

Experiment 1 found that children who were prompted to explain observations in a given domain needed fewer questions to complete a 20-questions task involving the same domain. Children who were presented with the same observations, but who were not prompted to explain, did no better than those who did not receive training. We also found that those who received structured observations (but no prompt to explain) performed no better than those who received no training at all. This suggests that *explaining* observations, rather than merely being exposed to them, is what drove children's improved performance.

Our results also revealed a developmental trend in the effect of explanation prompts on children's question-asking efficiency: statistically significant effects of explanation prompt were only found for older children, and the effects of prompt interacted with age. This is consistent with the results from Walker, Bonawitz, and Lombrozo (2017), who found that an explanation prompt increased 5-year-old children's preference for a simpler hypothesis (soil type over low water plus low light), but had no effect on 4-year-olds. They attribute the null effect to younger children's deficient prior knowledge concerning factors that affect plant growth. Similarly, the younger children in our task may have lacked crucial prerequisites to benefit from explanation, such as the ability to identify relevant shared features to construct good explanations during training. Consistent with this idea, we found that older children generated more accurate explanations during training, and that accuracy was marginally associated with more efficient performance in the 20-questions game. It could be that by generating accurate explanations, older (but not younger) children identified features that supported questions that went beyond the individual-object level, and that they thereby achieved more efficient performance on the task.

We also found that for children prompted to explain, subsequent questions involved a shift away from questions at the individual level. This is consistent with our proposal from the introduction, that a key effect of explanation could be to facilitate the identification of shared features across sets of items, resulting in questions that target more than one object.

In Experiment 2 we replicate these effects with children who span a greater age range, and with a more stringent control condition.

Experiment 2

Experiment 2 had two primary aims. First, we aimed to replicate the developmental pattern from Experiment 1, and the result that effects of explanation (for older children) manifest in a shift away from questions at the individual level. Participants from Experiment 2 therefore straddled our initial age range: we tested children who were 4- or 7 years of age (compared to the two age groups of 5- and 6-year-olds obtained by median split in Experiment 1).

Second, we aimed to address a potential concern from Experiment 1 by using a more stringent control condition. In Experiment 1, we discovered that the Training-with-explanation and Training-without-explanation conditions were not effectively matched in terms of time: children prompted to explain spent more time attending to the stimuli after they were introduced (Training-with-explanation: M = 48s; SD = 30s; Training-without-explanation: M = 25s; SD = 15s), $F(2,44) = 10.92$, $p = .002$, $\eta^2 = .21$. It is therefore possible that this additional time during training – and nothing specific to explanation – is the reason for older children's more efficient performance on the test trial. We find this alternative unlikely, given that children in the Training-without-explanation condition did not differ significantly from those in the No-training condition, despite a more dramatic difference in their training times (some versus none). Nonetheless, we thought it was worthwhile to replicate Experiment 1 with a more demanding control condition, where children were asked to point out the training materials *twice*, thereby drawing their attention to the relevant observations for a longer period of time.

Finally, Experiment 2 aimed to test an additional hypothesis for how explanation could affect subsequent information search: by helping children appreciate the hierarchical structure of the stimuli. To test this possibility, we included a binary sorting task modeled after Ruggeri et al. (2016). As we explain below, however, children's poor performance on this task limited its value.

Method

Participants. Participants were 50 4-year-old children (28 female, M = 53.3 months; SD = 3.9 months) and 48 7-year-old children (24 female, M = 88.0 months; SD = 4.1 months) recruited at local museums and preschools in Berlin, Germany. The children were native German speakers or fluent in German; they were predominantly white and belonged to various social classes. Six additional children were excluded from analyses due to experimental error (N = 4) or because they did not want to continue the experiment (N = 2). Written informed consent was obtained from participants' parents and the local ethical review board at the Max Planck Institute for Human Development approved the study protocol ("Cupcakes"). Children received a small gift to thank them for their participation.

Design and procedure. In Experiment 2, children were randomly assigned either to the Training-with-explanation or to the Training-without-explanation conditions. The procedure closely followed that of Experiment 1, with two crucial differences. First, children in the Training-without-explanation condition were asked *twice*, for each monster, to point to the treats that gave and did not give Seria/Menna/Polla/Nara a tummy ache. In this way, we made sure that children in this condition were not attending to the training materials for a shorter time than the children in the Training-with-explanation condition. Second, at the end of the main task, children performed a binary sorting task (as in Ruggeri et al., 2016) to determine whether they understood the hierarchical structure of the stimuli and were able to verbally label categories at each level. In the sorting task, children were given the 16 cards with the treats used for the test trial, and were asked to "divide the cards in two groups [higher level: cupcakes versus other treats]. They should be sorted so that all cards in one group are somehow similar to each other, and all cards in the other group are similar to each other. Also, the cards in one group should somehow differ from the cards in the other group. Can you do that?" Children were then asked, using similar instructions, to sort the cards within each of these two subgroups into two new piles (middle level: e.g., salty versus sweet treats among the non-cupcakes), and the third step required sorting the cards within each of these four subgroups into two new piles (lower level: e.g., candies versus chocolate). At each step, for both sorting tasks, children were asked to name each group they had sorted the objects into (i.e., "How do we call this group?"). When children, at any stage of the sorting task, did not organize the objects according to the expected hierarchical categorization (e.g., if they sorted them by color), the experimenter prompted the participant to sort the objects differently (i.e., "is there another way to sort the objects into two groups?"), without suggesting any specific way to do so. Thus to successfully perform the task, children had to identify the features needed to sort the given objects into groups at different hierarchical levels, beginning with features at the highest level.

An additional change was that the instructions and experimenter scripts were translated into German, and the experiment was conducted in German.

Results

We succeeded in eliminating the significant difference in time spent attending to the training stimuli across conditions: average time in the Training-with-explanation and Training-without-explanation conditions did not differ (Training-with-explanation: M = 80s; SD = 25s; Training-without-explanation: M = 76s; SD = 38s), p = .569.

Number of questions. We analyzed the number of questions required to complete the question-asking test trial as the dependent variable in a univariate ANOVA with condition as an independent variable, revealing no main effect of condition, but a main effect of age, $F(1,98) = 7.38$, $p = .008$, $\eta^2 = .06$, with 7-year-olds requiring fewer questions to complete the task (M = 7.67; SD = 3.32) than 4-year-olds (M = 9.82; SD =

4.72). The analysis also revealed an interaction between condition and age, $F(1,98) = 4.16$, $p = .044$, $\eta^2 = .04$. As can be seen in Figure 4, 7-year-olds needed fewer questions in the Training-with-explanation ($M = 6.65$; $SD = 2.55$) as compared to the Training-without-explanation condition ($M = 8.60$; $SD = 3.70$; $t(46) = 2.42$, $p = .020$), whereas the number of questions needed by 4-year-olds did not differ between conditions ($p = .299$).

Figure 4. Number of questions children needed to reach the solution, displayed by condition (Training-with-explanation and Training-without-explanation) and age group (4- and 7-year-olds). Bars represent one SEM in each direction.

Kinds of questions asked. As in Experiment 1, we analyzed the percentage of questions asked of each type as a function of condition and age group (see Table 3). These analyses revealed no main effect of condition on the percentage of questions at the lower ($p = .400$) or higher ($p = .693$) levels, nor were there main effects nor interactions

(p > .080). However, there were marginal effects of condition at the middle level, and significant effects at the individual-object level.

Table 3

Mean Percentage (and SDs) of Questions That Targeted The Lower, Middle, Higher Category Levels or Individual Objects in Experiment 2, Displayed By Condition (Training-with-explanation and Training-without-explanation) and Age Group (4- and 7-year-olds).

Age group	Condition	Individual-Object level	Lower level	Middle level	Higher level
4-year-olds	Training-with-explanation	50 (42)	11 (19)	31 (40)	8 (14)
	Training-without-explanation	53 (39)	15 (32)	17 (28)	14 (28)
7-year-olds	Training-with-explanation	39 (25)	22 (19)	20 (24)	19 (26)
	Training-without-explanation	69 (19)	10 (13)	11 (10)	9 (15)

For questions at the middle level, there was a marginal effect of condition, $F(1,94) = 3.58$, $p = .062$, $\eta^2 = .04$. Children asked a higher percentage of middle-level questions in the Training-with-explanation condition (M = 25%; SD = 33%), as compared to the Training-without-explanation condition (M = 15%; SD = 21%). We did not find a main effect of age (p = .135), nor an interaction (p = .662).

For questions at the individual-object level, there was a significant main effect of condition, $F(1,94) = 5.86$, $p = .017$, $\eta^2 = .06$, qualified by an interaction with age, $F(1,94) = 3.94$, $p = .050$, $\eta^2 = .04$. Overall, children asked a lower percentage of questions targeting only one treat in the Training-with-explanation condition (M = 45%; SD = 35%), as compared to the Training-without-explanation condition (M = 61%; SD = 32%; see Figure 5). However, two t-tests separated by age group confirmed that fewer of the 7-year-olds' questions in the Training-with-explanation condition targeted only one treat as compared to the Training-without-explanation condition, $t(48) = 4.48$, $p < .001$), whereas for 4-year-olds there was no difference between the Training-with-explanation and the Training-without-explanation conditions ($p = .795$). We did not find a main effect of age ($p = .769$).

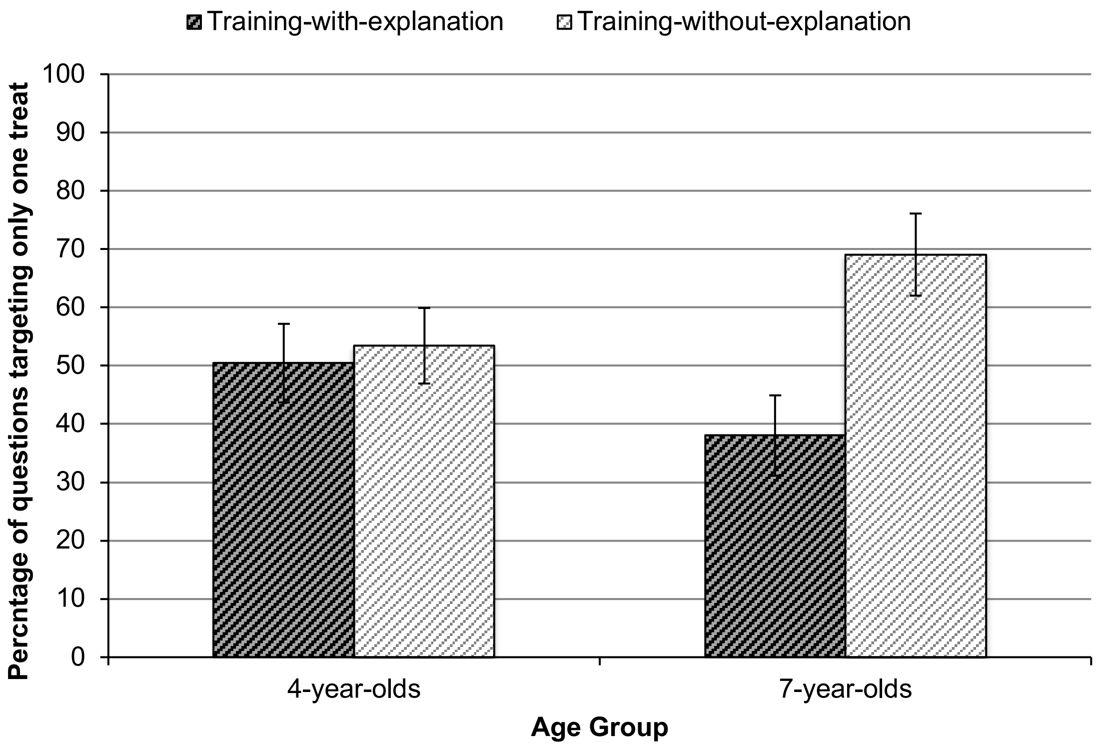

Figure 5. Percentage of questions children asked that targeted only one treat, displayed by condition (Training-with-explanation, Training-without-explanation and No-training) and age group (4- and 7-year-olds). Bars represent one SEM in each direction.

Final explanations. Although the percentage of children providing a correct final explanation was higher in the Training-with-explanation ($M = 30\%$) as compared to the Training-without-explanation condition ($M = 22\%$), a Chi-square test confirmed that the difference was not significant (p = .384). More 7-year-olds provided a correct final explanation (M = 40%) than 4-year-olds (M = 12%; p = .001).

Explanations during training. Explanations were coded as in Experiment 1, with 100% agreement. A univariate ANOVA showed that 7-year-olds provided more correct explanations during training (M = 1.83; SD = 1.56) than 4-year-olds (M = .38; SD = .92), $F(1,42) = 13.34$, $p = .001$, $\eta^2 = .25$. We also found a negative correlation between the number of correct explanations offered during training and the total number of questions needed to reach the solution (Pearson's r = -.426, p = .004). The correlation disappeared once we controlled for age, but there was a marginal effect even within the 7-year-old group alone (4-year-olds: r = -.226, p = .325; 7-year-olds: r = -.365, p = .095).

We also found, as in Experiment 1, that the number of questions needed to reach the solution negatively correlated with having mentioned "cupcakes" during training: those children who mentioned cupcakes on average needed 6.25 questions (SD = 2.67), whereas those who did not mention cupcakes required an average of 9.87 questions (SD = 4.45), $t(45) = -2.98$, $p = .005$).

Finally, there was a significant correlation between the number of correct explanations offered during training and the ability to provide a correct final explanation in the test phase: those children who provided a correct final explanation on average provided 2.31 correct explanations during training (SD = 1.70), whereas those who provided an incorrect final explanation on average provided 0.60 correct explanations during training (SD = 1.00), $t(41) = 4.12$, $p < .001$.

Sorting task. Very few children managed to successfully complete the sorting task without help from the experimenter (4-year-olds: 6%; 7-year-olds: 29%; $t(96)$ =

3.90, $p < .001$). In fact, many children were not able to complete the task at all, even *with* help from the experimenter (4-year-olds: 45%; 7-year-olds: 10%; $t(96) = -3.04$, $p = .003$). We therefore hesitate to draw strong conclusions from performance on the sorting task.[2]

Discussion of Experiment 2.

Experiment 2 replicated key results from Experiment 1. First, we again found significant effects of prompts to explain, despite matching the time on training across conditions. Second, the findings confirmed a strong developmental difference, whereby 7-year-olds, but not 4-year-olds, benefited from a prompt to explain. Third, our analyses once again revealed that older children who were prompted to explain asked different kinds of questions. We replicated the finding that prompts to explain led older children to ask fewer questions at the individual-object level, but did not find reliable differences in the level at which higher-level questions were asked.

The results from Experiment 2 shed light on the mechanisms underlying the effects of explanation on our younger and older participants. In particular, we find support for the idea that explanation helps older children identify shared features across items, and that this supports subsequent questions that go beyond the individual-object level. We also find support for the idea that younger children may have failed to benefit from a prompt to explain because they lacked the relevant knowledge or skill to identify shared features to constructs good explanations. During training, younger children produced fewer accurate explanations than older children, and this variation in accuracy

[2] Because performance on the sorting task was very poor, we are disinclined to draw conclusions from the null results that ensued: There was neither a main effect of experimental condition nor an interaction between age and condition on sorting performance (all *ps* > .332). We also failed to find a correlation between performance in the 20-questions game and performance in the sorting task (all *ps* > .157).

was associated with question-asking performance. Unfortunately, because performance in the sorting task was quite poor[3], we were unable to test the additional hypothesis that prompts to explain helped children appreciate the hierarchical structure of the stimuli. However, children's questions offer some evidence against this possibility: neither Experiment 1 nor 2 found that in shifting away from individual-object level questions, children uniquely favored the most efficient questions: those at the highest level of the hierarchy.

General Discussion

Across two experiments, we investigated whether prompting 4- to 7-year-old children to explain observations involving sets of objects prepares them to ask effective questions in a subsequent information-search task involving those objects. We found that prompts to explain can indeed improve the ability of 6-year-olds (Experiment 1) and 7-year-olds (Experiment 2), but not younger children, to ask efficient questions, and that it does so by facilitating the identification of features they can be used to ask questions that go beyond the individual-object level. These findings reveal that explaining not only plays a crucial role in learning by improving learning outcomes (Bonawitz, van Schijndel, Friel, & Schulz, 2012; Legare, 2012; Legare & Gelman, 2014; Roy & Chi, 2005; Siegler, 2002; Sobel & Sommerville, 2009; Wellman & Liu, 2007), but can also influence *how* children learn by preparing them to learn more efficiently.

The effects of our relatively simple explanation training are particularly striking in light of the fact that previous, more explicit attempts to improve children's question-asking strategies have met with only moderate success (e.g., Courage, 1989; Denney,

[3] This is surprising, considering that in Ruggeri et al. (2016) 84% of the 7-year-olds were able to complete an identical sorting task, only with different stimuli, without scaffolding.

1972). However, it is an open question whether alternative active prompts, such as asking children to compare the stimuli or to sort them into categories prior to the 20-questions game, would have had a similarly beneficial effect on children's performance. With adults, evidence suggests that general prompts to describe (Williams & Lombrozo, 2010) or to compare (Edwards, Williams, Gentner, & Lombrozo, 2018) do not yield equivalent effects to prompts to explain, but it may well be that explanation triggers *particular* descriptions or comparisons that support effective learning. In particular, there's evidence that prompting adults to explain in a category learning task leads them to compare *groups* of category members, and that this group-based comparison facilitates the discovery of shared and diagnostic features (Edwards, Williams, Gentner, & Lombrozo, 2018).

Our analyses of children's questions sheds light on what it is about explaining that affected subsequent performance on the hierarchical 20-questions task. In the introduction, we motivated the idea that explaining why some objects, as opposed to others, share a certain feature could encourage children to identify shared and diagnostic features for belonging to one set versus the other. Our data support this hypothesis: older children who were prompted to explain often succeeded in identifying such features during training, and this success was associated with later performance. Older children who were prompted to explain also succeeded in going beyond the individual-object level more reliably than other children. However, we did not find that children systematically focused on questions at the highest level of our hierarchical structure, and the hints that they shifted to the middle level in Experiment 1 were not replicated in Experiment 2. This suggests that the main effect of explanation was to help children identify features shared by several objects and thereby go beyond the object-level, rather than helping them appreciate and exploit the hierarchical structure of the task by selectively targeting the highest available level.

An interesting question for future research is whether our age-related findings are due to development per se, or instead reflect a more general effect of familiarity with the stimuli presented, or expertise. In our training task, we found that younger children were not very successful when it came to generating explanations, but other studies have found that 4- and 5-year-old children have at least some ability to generate good explanations (e.g., Bonawitz et al., 2012; Legare, 2012), and explanation prompts have been shown to have effects in children as young as 3 (Legare & Lombrozo, 2014; Walker et al., 2014). We speculate that with "easier" materials, e.g., objects or events children are familiar with, even 4- and 5-year-olds might show an influence of explanation prompts on information search, and that with more complex or novel materials, older children (and in some cases even adults) may not. Moreover, explanation prompts are unlikely to have much effect when familiarity or expertise make shared features salient at the outset.

Another open question concerns children's *flexibility* in question-asking, and whether flexibility might interact with explanation. In some cases, asking questions at the individual object level can actually be the most effective strategy (e.g., Ruggeri & Lombrozo, 2015; Ruggeri, Sim, & Xu, 2017). Imagine you are a good friend of the monster Loma, and you have observed her eating treats many times. You might realize that Loma often feels sick after having eaten a particular mint-flavored chocolate. In this case, if Loma gets sick again and you had to find out what made her sick, asking about cupcakes might not be the most effective strategy. It could make more sense to ask directly if it was the mint-favored chocolate. In our current task, children benefited from asking questions that went beyond the individual-object level. But recognizing that this is not always the case (see Ruggeri & Lombrozo, 2015) raises an intriguing question: might prompts to explain be *detrimental* to performance when an individual-object strategy is best? Or might explanation help children appreciate the structure of the task in a way that helps them cater their question-asking to the explained environment?

Finally, we acknowledge that our current stimuli and procedure are quite artificial. However, note that our primary aim was not to develop a realistic intervention for educational or everyday settings, but rather to use our intervention as a window onto question asking and explanation. In particular, this work takes a step towards understanding explanation and question asking as crucial components of an ongoing process of inquiry featuring the child as an active learner.

Acknowledgments

We would like to thank Mandana Mostofi, Susana Herrera, Minh-Thy Nguyen, Lara Bertram and Eva Kell for assistance in recruiting, data collection and coding. We would also like to thank all the children from the Laurence Hall of Science, in Berkeley, the Bay Area Discovery Museum, in Sausalito, and the Museum für Naturkunde, Labyrinth Museum and Forscherwelt Blossin, in Berlin for providing research space. This research was supported by a Seventh European Community Framework Programme Marie Curie International Outgoing Fellowship awarded to A. Ruggeri. T. Lombrozo was supported by a McDonnell Scholar Award on Understanding Cognition. F. Xu was supported by an NSF grant (SMA-1640816).

References

Berretty, P. M., Todd, P. M., and Martignon, L. (1999). Categorization by elimination: using few cues to choose, in *Simple Heuristics thatMake us Smart, eds G. Gigerenzer*, P. M. Todd, and The ABC Research Group (New York, NY: Oxford University Press), 235–254.

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*, 215–234.

Chi, M., Bassok, M., Lewis, M., Reimann, P. & Glaser, R. (1989). Self-explanations: how students study and use examples in learning to solve problems. *Cognitive Science, 13*, 145-182.

Chi, M., de Leeuw, N., Chiu, M.H., LaVancher, V. (1994). Eliciting self-explanations improves understanding. *Cognitive Science, 18*, 439-477.

Chin-Parker, S. & Cantelon, J. (2017). Contrastive constraints guide explanation-based category learning. *Cognitive Science, 41*(6), 1645-1655.

Courage, M. L. (1989). Children's Inquiry Strategies in Referential Communication and in the Game of Twenty Questions. *Child Development*, *60*(4), 877. https://doi.org/10.2307/1131029

Crowley, K., & Siegler, R. S. (1999). Explanation and Generalization in Young Children's Strategy Learning. *Child Development*, *70*(2), 304–316. https://doi.org/10.1111/1467-8624.00023

Denney, D. (1972). Modeling and eliciting effects upon conceptual strategies. *Child Development*, 43 (3), 810–823.

Denney, D., Denney, N., & Ziobrowski, M. (1973). Alterations in the information-processing strategies of young children following observation of adult models. *Developmental Psychology*, 8 (2), 202–208.

Denney, N., & Turner, M. (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.

Edwards, B. J., Williams, J. J., Lombrozo, T., & Gentner, D. (under review). Explanation recruits comparison: Insights from a category-learning task.

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

Hamilton, G. C., Sanders, A., Strange, G. S., & Trott, A. T. (2003). *Emergency Medicine: An Approach to Clinical Problem Solving*, *2nd Edn.* Philadelphia, PA: 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. http://doi.org/10.1016/0022-0965(82)90015-7

Kon, E., & Lombrozo, T. (2018). Seeking Ideal Explanations in a Non-Ideal World. T. T. Rogers, M. Rau, X. Zhu, & C. W. Kalish (Ed.), Proceedings of the 40th Annual Conference of the Cognitive Science Society (1939-1944) . Austin, TX: Cognitive Science Society.

Legare, C. H. (2012). Exploring explanation: Explaining inconsistent information guides hypothesis-testing behavior in young children. *Child Development, 83*, 173–185.

Legare, C. H., & Gelman, S. A. (2014). Examining explanatory biases in young children's biological reasoning. *Journal of Cognition and Development, 15*, 287–303.

Legare, C. H., & Lombrozo, T. (2014). Selective effects of explanation on learning during early childhood. *Journal of Experimental Child Psychology*, *126*, 198–212. doi:10.1016/j.jecp.2014.03.001

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*, 63–76. doi:10.1016/j.jecp.2012.07.002

Lombrozo, T. (2006). The structure and function of explanations. *Trends in Cognitive Sciences*, *10*(10), 464–70. https://doi.org/10.1016/j.tics.2006.08.004

Lombrozo, T. (2012). Explanation and abductive inference. In K.J. Holyoak and R.G. Morrison (Eds.), *Oxford Handbook of Thinking and Reasoning* (pp. 260-276), Oxford, UK: Oxford University Press.

Lombrozo, T. (2016). Explanatory Preferences Shape Learning and Inference. *Trends in Cognitive Sciences*, *20*(10), 748–759. doi.org/10.1016/j.tics.2016.08.001

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

Mosher, F. A., & Hornsby, J. R. (1966). On asking questions. *Studies in Cognitive Growth*. New York, NY: 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. doi:10.1016/j.cognition.2013.09.007

Roy, M., & Chi, M. (2005). The self-explanation principle in multimedia learning. In R. Mayer (Ed.), *Cambridge handbook of multimedia learning* (pp. 271–286). Cambridge, UK: Cambridge University Press.

Rothe, A., Lake, B. M., & Gureckis, T. M. (2018). Do People Ask Good Questions? *Computational Brain & Behavior*, *1*(1), 69–89. http://doi.org/10.1007/s42113-018-0005-5

Ruggeri, A., & Feufel, M. a. (2015). How basic-level objects facilitate question-asking in a categorization task. *Frontiers in Psychology*, *6*(July), 1–13. doi:10.3389/fpsyg.2015.00918

Ruggeri, A., & Lombrozo, T. (2015). Children adapt their questions to achieve efficient search. *Cognition*, *143*, 203–216. doi:10.1016/j.cognition.2015.07.004

Ruggeri, A., Lombrozo, T., Griffiths, T. L., & Xu, F. (2016). Sources of developmental change in the efficiency of information search. *Developmental Psychology*, *52*(12), 2159–2173. https://doi.org/10.1037/dev0000240

Ruggeri, A., Sim, Z. L., & Xu, F. (2017). "Why is Toma late to school again?" Preschoolers identify the most informative questions. *Developmental Psychology*, *53*(9), 1620–1632. https://doi.org/10.1037/dev0000340

Ruggeri, A., Walker, C., Lombrozo, T., & Gopnik, A. *Strategic use of questions to solve a causal inference task in early childhood.* Manuscript submitted for publication.

Schult, C. A., & Wellman, H. M. (1997). Explaining human movements and actions: Children's understanding of the limits of psychological explanation. *Cognition*, *62*(3), 291–324. https://doi.org/10.1016/S0010-0277(96)00786-X

Siegler, R. S. (2002). Microgenetic studies of self-explanation. In N. Granott & J. Parziale (Eds.), *Microdevelopment: Transition processes in development and learning* (pp. 31–58). Cambridge, UK: Cambridge University Press.

Sobel, D. M., & Sommerville, J. A. (2009). Rationales in children's causal learning from others' actions. *Cognitive Development, 24,* 70–79.

Walker, C.M., Bonawitz, E., & Lombrozo, T. (2017). Effects of explaining on children's preference for simpler hypotheses. Invited paper for *Special Issue: The process of explanation* in *Psychonomic Bulletin & Review.* doi: 10.3758/s13423-016-1144-0.

Walker, C.M. & Lombrozo, T. (2017). Explaining the moral story. Invited paper for *Special Issue: Moral Learning* in *Cognition, 167,* 266-281.

Walker, C. M., Lombrozo, T., Legare, C. H., & Gopnik, A. (2014). Explaining prompts children to privilege inductively rich properties. *Cognition*. https://doi.org/10.1016/j.cognition.2014.07.008

Walker, C. M., Lombrozo, T., Williams, J. J., Rafferty, A. N., & Gopnik, A. (2017). Explaining Constrains Causal Learning in Childhood. *Child Development*, *88*(1). https://doi.org/10.1111/cdev.12590

Wellman, H. M. (2011). Reinvigorating explanations for the study of early cognitive development. *Child Development Perspectives*, *5*(1), 33-38.

Wellman, H. M., & Liu, D. (2007). Causal reasoning as informed by the early development of explanations. In A. Gopnik & L. Schulz (Eds.), *Causal learning:*

Psychology, philosophy, and computation (pp. 261–279). Oxford, UK: Oxford University Press.

Williams, J. J., & Lombrozo, T. (2010). The role of explanation in discovery and generalization: Evidence from category learning. *Cognitive Science, 34*, 776–806.

Williams, J. J., Lombrozo, T., & Rehder, B. (2013). The hazards of explanation: Overgeneralization in the face of exceptions. *Journal of Experimental Psychology: General, 142*(4), 1006-1014. doi: 10.1037/ a0030996.

Wilkenfeld, D. A. & Lombrozo, T. (2015). Inference to the best explanation (IBE) versus explaining for the best inference (EBI). *Science and Education*. 1-19. doi: 10.1007/s11191-015-9784-4

Supplementary Materials

The Database containing the data for the two Experiments included in this manuscript is archived in the Open Science Framework repository: https://osf.io/ptqn8/files/ (Database: Ruggeri, Xu, & Lombrozo, 2019).