*Child Development*, 2026, 1–16
https://doi.org/10.1093/chidev/aacag040
Published: May 25, 2026
**Empirical Article**

# Comparison helps children form broad explanations

**Yinyuan Sean Zheng[1,*] and Dedre Gentner[1]**

[1]Department of Psychology, Northwestern University, Evanston, IL, United States
*Corresponding author: Department of Psychology, Northwestern University, 2029 Sheridan Road, Evanston, IL 60208, United States. Email: zhengyinyuan@u.northwestern.edu

## Abstract

**Children gain insight from trying to explain events, in part because self-explaining leads children to focus on broad general patterns. This study examines how children come to discover these broad patterns. We propose that generating explanations invokes comparison, which reveals commonalities that feed into a broad explanation. In three experiments, 4- to 6-year-olds ($N = 138$; 61 girls; $M_{age} = 5;30$; predominantly White; data collected 2023–2024) explained a pattern of outcomes that could result either from a single broad cause or from several specific causes. Children were more likely to arrive at the broad explanation when comparison was available than when it was not (E1: $d = 1.67$; E2: $d = 1.08$). The degree of required comparison support lessoned with age. Comparison and self-explanation combined to generate satisfactory explanations (E3: $d = -0.91$).**

**Keywords** comparison, self-explanation, development

## Lay summary

Children often gain insight from trying to explain events. We suggest that this is because the effort to explain leads children (and adults) to seek comparisons across the items to be explained. This process of comparison can reveal important commonalities, which leads children to focus on broad general patterns instead of local details. This claim found support in three experiments. 5- and 6-year-old children were more likely to arrive at a broad explanation when comparison was readily available than when it was not; and 4-year olds were more likely to arrive at a broad explanation when comparison was encouraged than when it was not. Explanation and comparison may work together to spur learning and curiosity.

Children are avid learners. They seek understanding both by asking questions like "Why is the sky blue?" and by framing their own explorations—for example, experimenting with towers of blocks to understand what makes them fall. Developmental research has revealed key processes that support children's learning, including self-explanation, causal reasoning, and analogical comparison. *Self-explanation*—the phenomenon whereby a child (or adult) tries to explain an event or situation to themselves—has been increasingly studied of late. There is considerable evidence that engaging in self-explanation promotes learning, both in children (Bisra et al., 2018; Chi et al., 1994; Rittle-Johnson, 2006; Rittle-Johnson et al., 2017; Walker et al., 2014; Walker & Lombrozo, 2017) and adults (Chi et al., 1989; Lombrozo, 2016; Muldner et al., 2014; Nokes-Malach et al., 2013). For example, in her pioneering studies, Chi and colleagues asked students to think aloud while reading a physics text (Chi et al., 1989). The results showed that students who actively explained the text to themselves as they read did better on later tests of comprehension and inference.

One reason that generating explanations is beneficial to learning is that people tend to prefer explanations that invoke broad, generalizable regularities (Legare et al., 2017; Lombrozo, 2006; Walker & Lombrozo, 2017; Williams & Lombrozo, 2010). This explanatory preference naturally benefits transfer by focusing learners on generalizations that hold over a broad set of phenomena (Busch et al., 2018; Lombrozo, 2016; Walker et al., 2014). This raises a key question—*how* does explaining lead learners to discover broad patterns? We suggest that one key process invoked during explanation is analogical comparison over instances of the phenomenon to be explained. Comparison then fosters abstracting common structure from the examples compared. Before presenting evidence, we first briefly review the literature on the learning benefits of explanation generation, with a focus on development.

## The learning benefits of explanation generation

In their classic studies, Chi and colleagues showed that engaging in self-explanation can help students grasp difficult science

**Handling Editor:** Dave Sobel
**Editor:** Adam Winsler
**Received:** July 28, 2025. **Revised:** February 17, 2026. **Accepted:** February 23, 2026

concepts (Chi, 2000; Chi et al., 1989, 1994). For example, Chi et al (1994) asked eighth graders either to self-explain a passage on the human circulatory system or to read the passage twice. The self-explanation group showed a deeper understanding of the topic than those who simply read the same text twice. Those who self-explained were more likely to elaborate on the material and to integrate the text content with their prior knowledge and were better able to answer inference questions (Chi et al., 1994).

Since then, much research has demonstrated the benefit of generating explanations to oneself. For example, Legare and Lombrozo (2014) showed 3- to 5-year-old children a set of gears that could be arranged to turn a fan. Half the children were asked to explain how it worked, while the other half merely observed the demonstration. During testing, children were first asked to repair the same machine that was missing a gear and then to construct a new one. Children in the explanation group were more successful in both fixing the machine and constructing their own machines than those in the observation group. This suggests that explaining can foster children's discovery of nonobvious causal patterns (Busch et al., 2018; Callanan & Oakes, 1992; Keil, 2006; Legare et al., 2017; Legare & Lombrozo, 2014; Lombrozo, 2006; Walker et al., 2014). Interestingly, in Legare and Lombrozo's study, 3- and 4-year olds (though not 5-year olds) showed worse perceptual memory of the gear colors when being asked to explain. Walker et al. (2014) replicated and extended this finding, showing that this tendency occurred even when children failed to generate a correct explanation. The authors concluded that the attempt to explain was sufficient to trigger the bias to attend to causal patterns at the expense of local properties. Research has also shown that generating explanations can benefit relational reasoning. Brockbank et al. (2023) presented 5- and 6-year olds a challenging analogy task, the Relational Match-to-Sample (RMTS) task with an object competitor (Christie & Gentner, 2007; Kroupin & Carey, 2021; Kroupin & Carey, 2021). In this task, children see pairs of identical shapes (AA) and are asked to choose which option is more similar, a relational match (BB) or a partial object match (AC). Past findings suggest that even 8.5-year old children do not spontaneously favor the relational match (Christie & Gentner, 2007), consistent with research showing a strong attraction toward object similarity in children (Gentner & Toupin, 1986). Brockbank et al. modified the procedure so that the experimenter demonstrated that the relational match should go with the target. Half the children were asked to explain why this was right; the other half simply described what happened. When children were asked to do the RMTS task themselves, those who had explained were more likely to respond relationally than those who had described the events.

According to Lombrozo's (2016; William and Lombrozo, 2010) subsumptive constraint account, one reason that explanation-generation benefits learning is that people prefer explanations that characterize things in terms of broad, general patterns. Thus, a key property of explanation is to favor generalization—that is, to discover that a specific event can be subsumed under a general pattern that also accounts for other cases. People find such explanations informative and therefore satisfying. Consequently, when constructing explanations themselves, people search for an explanation that reveals a broader pattern. In short, this view predicts that the act of explaining tends to lead learners to seek broad regularities rather than context-specific accounts. This explanatory preference has important implications for how we reason and learn (Lombrozo, 2006, 2016; Lombrozo & Carey, 2006).

To test this claim, Williams and Lombrozo (2010) presented adults with two groups of artificial robots. The categories were designed such that most of the members within each category shared a highly salient feature (body shape), but all members in a group shared a less salient property (foot shape). Specifically, the four "Glorps" all shared the subtle feature of pointy feet but only three of them had a round body; in contrast, the four "Drents" all shared the subtle feature of flat feet but only three had a square body. Thus, body shape constituted an imperfect, albeit salient, rule with a 75% success rate, whereas foot shape was a less obvious 100% rule that best divided the robots. When people were asked to explain why each robot belonged to its category, they were more likely to identify the 100% rule than were those who simply described the robots. These findings support the claim that engaging in explanation seeking drives a search for unifying broad patterns. Similar results have been found for children (Walker et al., 2017; Walker & Lombrozo, 2017). Given the effectiveness of self-explanation in learning, it is critical to understand what mechanisms support its success.

## The role of comparison in explanation

How do people arrive at a broad explanation? Sometimes the relevant knowledge can simply be retrieved—for example, when applying a familiar conceptual framework to explain new cases. But often a broad pattern needs to be identified on the spot. We suggest that analogical comparison is one key mechanism that allows learners to arrive at broad patterns. The idea that comparison is implicated in generating explanations has received some support (Chin-Parker & Bradner, 2017; Hilton & Slugoski, 1986; Hoyos & Gentner, 2017; Landy & Hummel, 2010; Sidney et al., 2015). According to structure-mapping theory, comparison entails a process of structural alignment in which like relations are aligned and objects are placed into correspondence based on having like roles within the common relational structure (Forbus et al., 2017; Gentner, 1983, 2010; Gentner & Forbus, 2025). There is considerable evidence that structure-mapping renders common relational structure more salient (Christie & Gentner, 2010; Gentner, 2003, 2010; Gentner & Namy, 1999; Gick & Holyoak, 1983; Loewenstein et al., 2003; Markman & Gentner, 1993).

Thus, we propose that comparison contributes to explanation's effects by revealing potentially relevant commonalities and fostering them into a more general representation. This fosters knowledge abstraction, in which a general idea subsumes specific instances (Chen, 1999; Dixon & Bangert, 2004; Gick & Holyoak, 1983; Markman & Gentner, 1993; Thompson & Opfer, 2010). The abstraction may already be known to the learner, but it need not be—structural alignment can reveal novel relational commonalities that were not obvious in either of the separate items beforehand (Christie & Gentner, 2010; Gentner & Hoyos, 2017; Gentner & Namy, 1999). In sum, we propose a general pattern in which (1) generating explanations invokes comparison and (2) the comparison process highlights potentially important commonalities (and sometimes differences) that provide the basis for broad explanations.

A study by Edwards et al. (2019) provided evidence for this claim. The researchers presented people with the two robot categories used by Williams and Lombrozo (2010). Again, the robots could be differentiated using a 75% accurate salient feature (i.e., body shape) or a 100% accurate, less obvious property (foot shape). Half the participants were asked to explain why a robot (or

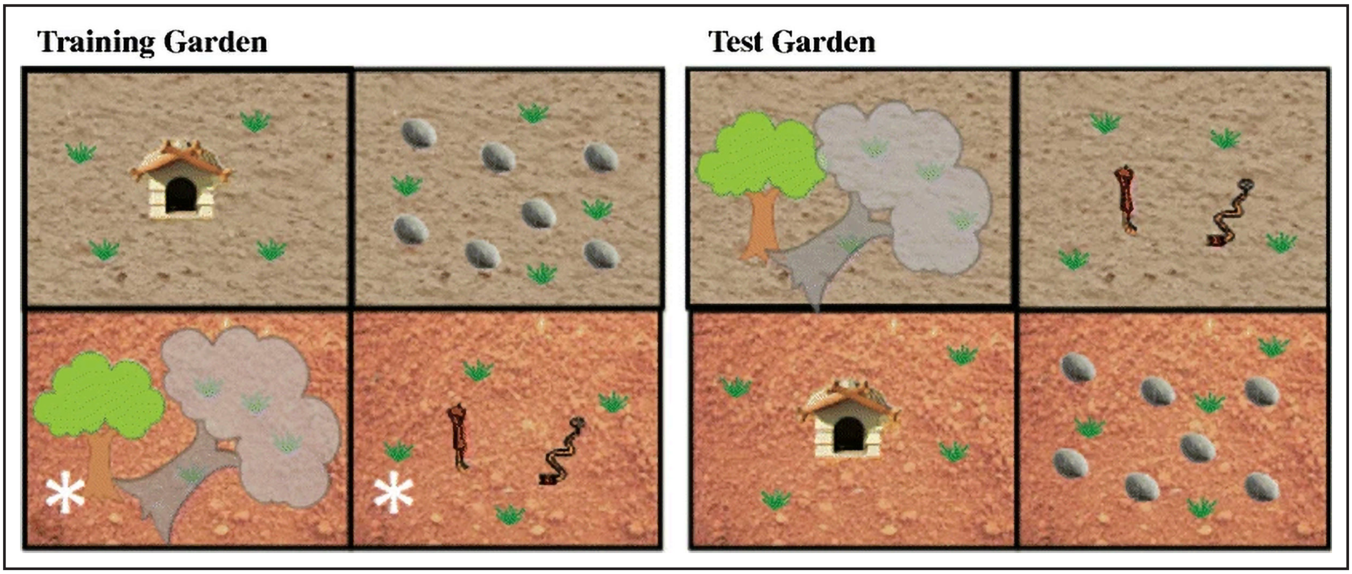


**Figure 1** Training (left) and test (right) gardens used in Walker et al. (2017) and in the current experiment. Both training gardens that grow healthy carrots have brown soil (top row), in which one contains a doghouse and the other contains rocks; both training gardens that grow sick carrots have red soil (bottom row), in which one contains a shady tree and the other contains a broken sprinkler. For the test gardens, the combinations of objects and soil types are reversed. The garden materials are used with permission from Walker et al. (2017).

robots) belonged to one category or the other. The other half were asked to compare pairs or groups of robots, either within-category or between-category or both. The results underlined the effectiveness of explanation generation: participants in the explanation conditions were more likely to discover the 100% rule than those in the comparison conditions. However, participants' self-reports revealed a telling pattern. As expected, people in the explanation condition reported engaging in more explanation than did those in the comparison condition. But strikingly, people in the explanation condition also reported engaging in more comparison than did those in the comparison condition. Further, the results of Experiment 3 showed that the success of the explanation instructions on rule discovery was partially mediated by the rate of self-reported group comparisons. Edwards et al. concluded that the prompts to explain led people to engage in spontaneous comparison, which then contributed to the effectiveness of explanation in identifying a broad regularity. Here we ask whether this pattern also occurs among children.

## Current study

This work tests the above claims in three experiments. Specifically, we asked whether comparison processes contribute to the effect of explanation generation on children's ability to discover broad regularities. To do this, we adapted a study paradigm by Walker et al. (2017). They showed that explaining can lead children to focus on broad causal patterns. We first describe Walker et al.'s study and then discuss a possible role for comparison.

Walker et al. (2017) tested the effect of explaining on children's causal reasoning by directly contrasting broad patterns versus specific details. They showed 4-, 5-, and 6-year olds four training gardens (Figure 1, left). Two of these had red soil, and two had brown soil. Each garden also contained a unique salient feature, such as a broken sprinkler or rocks. Children were told that some gardens grew sick carrots, an outcome that could be attributed to either a single cause or two independent causes. Children were shown that the two red soil gardens grew sick carrots. This was consistent with two hypotheses: a broad, common cause (i.e., both sick gardens had red soils) or a context-bound hypothesis that posits two separate causes (one had a broken sprinkler, so it lacked water, and the other had a shady tree, so it lacked sunlight). Half the children were asked to explain why the two carrots were sick, while the other half were simply asked to describe what they saw. The key question was whether explaining would lead children to prefer the broad (soil-based) hypothesis over the context-bound hypothesis that focused on garden-specific explanations such as "too shady" or "broken sprinkler."

To test children's preferred explanation, Walker et al. showed children four test gardens that reversed the combinations of soil types and specific objects (Figure 1, right). Children were asked to predict whether each garden grew healthy or sick carrots. If children endorsed the broad hypothesis, they should predict that the two red soil gardens (now containing the doghouse and rocks) would grow sick carrots, and the two brown soil gardens (now containing the shady tree and broken sprinkler) would grow healthy carrots. Their predictions would be the opposite if they endorsed the context-bound hypothesis.

The results showed an interesting developmental trend. Regardless of condition, 4-year olds did not consistently endorse either the broad or the context-bound hypotheses, while 6-year olds always endorsed the broad hypothesis. However, the 5-year olds behaved differently in the two conditions. In the *describe* condition, they were equally likely to choose the broad or context-bound hypotheses; but in the *explain* condition, they showed a clear preference for the broad hypothesis. As the authors suggested, this developmental shift could have resulted from gains in causal knowledge. The 4-year olds might lack sufficient domain knowledge of plant growth to think about the soil, while the 6-year-olds might know enough to favor the soil-type hypothesis even when not being prompted to explain. Critically, for the 5-year olds, the specific

**Table 1** Key design features in Walker et al. (2017) and in the current study.

| Condition | Walker et al. (4-, 5-, and 6-year olds) Explain | Walker et al. Describe | Experiment 1 (5- and 6-year olds) Explanation + Comparison Available | Experiment 1 Explanation + Comparison Difficult |
|---|---|---|---|---|
| Training phase | Training gardens shown simultaneously<br>Two healthy gardens introduced first, followed by two sick gardens | | Training gardens shown sequentially<br>Alternating between healthy and sick gardens | |
| Initial question | "Why do you think these two carrots are sick?" | "Were these two carrots healthy or sick?" | "Why do you think some carrots are sick?" | |
| Memory phase | N/A | | Gardens remain visible after shown | Gardens removed after shown |
| Test phase | Test gardens shown simultaneously<br>Training gardens visible on the side | | Test gardens shown sequentially<br>Training gardens visible on the side | Training gardens not visible |
| Final explanation | N/A | | "Why do you think these carrots are sick?" | |

| Condition | Experiment 2 (4-year olds) Explanation + Comparison Available (same as E1) | Experiment 2 Explanation + Comparison Prompted | Experiment 3 (5- and 6-year olds) Comparison-Only (no initial explanation) |
|---|---|---|---|
| Training phase | Training gardens shown sequentially<br>Alternating between healthy and sick gardens | Training gardens shown simultaneously<br>(1) Two healthy gardens introduced first, followed by two sick gardens (2) further prompted for explicit comparison | Training gardens shown sequentially<br>Alternating between healthy and sick gardens |
| Initial question | "Why do you think some carrots are sick?" | | N/A |
| Memory phase | Gardens remain visible after shown | | Gardens remain visible after shown |
| Test phase | Test gardens shown sequentially<br>Training gardens visible on the side | | Test gardens shown sequentially<br>Training gardens visible on the side |
| Final explanation | "Why do you think these carrots are sick?" | | "Why do you think these carrots are sick?" |

causes—lack of water and lack of sunlight—were often highly salient at first, but explaining allowed them to overcome them and consider the broad soil type alternative more thoroughly.

These findings underscore the value of explanation generation in prompting children to arrive at broad causal patterns. But the question we raise is, to what extent did the learning benefits of explanation stem from the availability of relevant comparisons? For example, the four training gardens were displayed together and remained present throughout the study. There is considerable evidence that children are more likely to compare things when they are simultaneously present than when they occur in sequence (Alfieri et al., 2013; Begolli & Richland, 2016; Christie & Gentner, 2010; Gentner & Hoyos, 2017). Thus, seeing the four gardens together could have allowed children to compare them, facilitating children's discovery of the common soil type.

In Experiment 1, we tested this possibility by repeating Walker et al.'s (2017) study with slight alterations designed to make comparison more difficult. We used the same materials and general procedure as did Walker et al. and the same age group (5- and 6-year-old children). One condition (Explanation + Comparison) was a close replication of Walker et al.'s Explain condition, with a few specific changes as noted below (see also Table 1). The other condition (Explanation Only) was matched to the Explanation + Comparison condition except for one critical difference: we minimized children's opportunity to compare. Specifically, for both conditions, the training gardens were shown one at a time (unlike Walker et al., who showed all gardens simultaneously). Children were told which gardens had healthy carrots and which had sick carrots, and all were asked to explain why some carrots were sick. We added a memory check to rule out concerns about encoding errors or memory loss. The critical opportunity for comparison was manipulated during the memory check. Children were shown the gardens one by one and asked to remember whether the garden had sick carrots or healthy carrots. Critically, for the Explanation + Comparison group, each garden remained in view after the memory question and stayed visible for the rest of the study (as in Walker et al.'s study). For the Explanation-Only group, the gardens were removed from sight after each question. Therefore, only the Explanation + Comparison group had access to readily available comparisons. All children then completed the Prediction task as in Walker et al.'s study. We also added a final explanation task, in which all children explained why they made their predictions. Table 1 summarizes the key design similarities and differences between the current study and Walker et al.'s (2017) study.

Our main prediction is that explainers who have access to readily available comparisons will be more likely to identify the broad pattern of common soil type—as reflected in their predictions and final explanations—than those for whom comparisons are made difficult.

# Experiment 1

## Method

### Participants

Fifty-two 5- and 6-year-old children were recruited for the study ($M = 5;11$ years, range = 5;00 to 6;11 years; 20 girls; 63% White, 21% Asian, 4% Black, 4% Hispanic, 8% Mixed; data collected 2023–2024): 26 in the Explanation + Comparison Available [Comp-Available] condition and 26 in the Explanation-Only [Comp-Difficult] condition. Families were recruited through a psychology department database from a midwestern city in the United States with medium-to-high average socioeconomic status. Families were compensated for participation. Assuming a power level of 80% and an alpha level of 95%, a total of 52 subjects were required for the independent samples $t$-test to detect an effect size of 0.8. One participant was excluded due to memory failure (see below).

### Materials

Two sets of four gardens were used (Figure 1). Each garden was printed on a 5.2 cm × 3.5 cm laminated paper. In both sets, the four gardens each contained a unique feature: a doghouse, rocks, a broken sprinkler, and a shady tree. Critically, the training set paired the doghouse and the rocks with brown soil, and the sprinkler and the tree with red soil (Figure 1, left), while the test set reversed this pairing (Figure 1, right). Green leaf clumps were used to indicate the carrot locations. The leaves were identical for sick and healthy carrots, so the child had to remember which gardens were sick and which were healthy. There were also a laminated picture of Mr. Farmer and two laminated pictures of carrots, one healthy and one sick. The stimuli, data, and analytic code for this study are available at https://osf.io/5zd36.

### Procedure

After obtaining parental consent and the child's verbal assent, the experimenter brought the child into a quiet room. The experiment consisted of three phases—Training, Memory, and Test. The Training phase was identical between conditions. The critical manipulation of comparison opportunity occurred at the Memory and Test phases. The study took 10 minutes.

*Training*

The experimenter first showed the child the picture of Mr. Farmer. They were told that he grows carrots, but that in some of his gardens, the carrots are sick. Children were invited to "…help Mr. Farmer figure out what makes the carrots healthy or sick." All children were then introduced to the four training gardens sequentially. For each, they were told, "In this garden, there is an X, and the soil is Y." (The X is "a doghouse," "rocks," "a broken sprinkler," or "a huge tree and its shadow," and the Y is "brown" or "red," respectively). The experimenter explicitly pointed to the objects and the soil while labeling these features. Next, the experimenter said, "Let's pick a carrot to see if the carrots in this garden are healthy or sick. Which one shall we pick?" After the child made a choice, the experimenter placed a healthy or sick carrot picture underneath a green leaf clump and said, "The carrots in this garden are healthy" for each of the two brown soil gardens, or "The carrots in this garden are sick" for each of the two red soil gardens. The experimenter then removed the garden picture before laying down the next one, so that the child could only see one garden at a time. The order of the gardens was fixed and alternated between healthy and sick gardens, so as to make the comparison less likely across gardens of the same soil type.

At the last garden, the child was asked for an *initial explanation* before the picture was taken away, "Mr. Farmer really wants to know what makes carrots sick. Why do you think some carrots are sick?" At this point, both hypotheses were compatible with the gardens—one based on the individual features (the context-bound hypothesis) and the other based on the common soil type (the broad hypothesis).

To summarize, we altered Walker et al.'s training procedure in two ways to minimize children's tendency to compare: (1) we showed gardens sequentially, whereas Walker et al. showed them simultaneously and (2) we alternated between healthy and sick gardens, whereas Walker et al. introduced the two healthy gardens together, followed by the two sick ones. One other change is that we used pictures of healthy carrots as well as sick carrots (unlike Walker et al., who only showed pictures of sick carrots). This was done to allow feedback when prompting children to check both healthy and sick gardens.

*Memory*

All children were again shown the four training gardens, in the same order as before. For each, children were asked, "Do you remember this garden? Does it grow healthy carrots or sick carrots?" Wrong answers were corrected. The exclusion criterion was a priori set to be more than one error. Overall, the children's memory was good. For the Comp-Available condition, four participants each made a single error; for the Comp-Difficult condition, one made an error and another made three (and was therefore excluded).

The two conditions differed in one specific way. In the Comp-Available condition, each garden was moved to one side of the table after the memory question and remained in view. In the Comp-Difficult condition, each garden was again taken away before laying down the next one. Therefore, by the end of the Memory phase, only the Comp-Available group could view all the gardens simultaneously, and the gardens stayed visible during the subsequent tasks. This is comparable to Walker et al.'s procedure in the Explain condition, in which children went directly from training to test, with the gardens remaining in view.

*Prediction task*

As in Walker et al.'s study, the four test gardens reversed the association between soil type and concrete details. In the Prediction Test, the experimenter randomly showed the four test gardens one at a time. For each one, they asked, "This is one of Mr. Farmer's carrot gardens. If I pick a carrot from this garden, do you think it will be healthy or sick?" The test garden was taken away after each response. This was done to make comparison difficult and differed from Walker et al.'s procedure, in which all four test gardens were shown together. The procedure was the same for both conditions, except that only the Comp-Available condition had access to the training gardens on the side.

Last, the experimenter showed each child all the test gardens that the child had predicted to grow sick carrots and asked, "Why do you think they grow sick carrots?" If all test gardens were predicted to be healthy, the experimenter asked, "Why do you think they grow healthy carrots?" These responses constituted the child's *final explanation*.

# Scoring and coding

## Prediction task

For each test garden, a child was given one point if the response was in accord with the broad hypothesis (i.e., that the two red soil gardens would produce sick carrots, and the two brown soil gardens would produce healthy ones) and zero otherwise. The max total score was 4. As in Walker et al. (2017), we classified children's responses into six categories based on the patterns—"All Soil" if the score was 4, indicating all predictions were aligned with the broad hypothesis; "All Context-Bound" if the score was 0; "Sick Bias"/"Healthy Bias" if children predicted all gardens to be sick/healthy; "Sick + 1" if children predicted sickness for the two red soil gardens and one more brown soil garden; and "Others" for all other responses.

## Initial and final explanations

We coded children's explanations according to their causal content. Four coding categories were specified—whether an explanation mentioned (1) soil, (2) lack of water (or a broken sprinkler), (3) lack of sunlight (or a shady tree), or (4) any other reason. For each category, the scoring was binary: "present" if the explanation mentioned the corresponding cause or "absent" otherwise. In reporting the results, we retained the "Soil" (the broad explanation) and "Other" categories but combined the "Water" and "Sunlight" into one category "Individual Feature," since both invoked context-specific features of the sick training gardens. This category was scored as "present" if either or both of the featural causes were present. One explanation instance could be scored in multiple categories (see Table 2).

All instances of verbal explanation (including those from Experiments 2 and 3) were compiled together, with participant and condition information hidden. The first author coded all the verbal responses, and two research assistants, who were not told the hypotheses, each coded half of the responses. Cohen's Kappa was calculated for interrater agreement: (1) "Soil", $k = 0.87$; (2) [lack of] "Water", $k = 0.92$; (3) [lack of] "Sunlight", $k = 0.96$; (4) "Other Reasons", $k = 0.80$. The average Kappa agreement was 0.89. Given the high agreement, we opted to use the first author's scores as the final codes.

For the analysis, we focused on two measures: the relative frequency of mentioning each cause ("Soil type," "Individual Feature," and "Other") and the frequency of mentioning soil type as the *sole* cause. The former reveals which hypothesis children preferred, and the latter reveals the extent to which children focused on the broad hypothesis of soil type.

# Experiment 1 results

All analyses were conducted in R (Version 4.4.1; R Core Team, 2024). For the verbal analysis on the explanation codes, we used a Chi-square test whenever all cells had count frequencies above or equal to 5 and a Fisher's exact test whenever any cell had a low count (<5).

For the prediction task, we first carried out an omnibus ANOVA including condition (Comp-Available or Comp-Difficult), age group (5- or 6-year old), and their interaction as independent variables. The results showed that neither age, $F(1, 47) = 1.66$, $p = .2$, nor its interaction with condition, $F(1, 47) = 5.5$, $p = .08$, was significant. Importantly, the two conditions differed significantly on the number of choices in favor of the broad hypothesis, $F(1, 47) = 37.84$, $p < .001$. This simple effect was followed up by an independent $t$-test as reported below. Given the reported age trend in Walker et al., we also examined each age group separately. The results showed similar patterns between 5- and 6-year olds on both the prediction task and the verbal explanations. These results are reported in the supplementary materials.

## Prediction task

(Figure 2) A $t$-test revealed a significant effect of condition, $t(45) = 5.94$, $p < .001$, $d = 1.67$. Children in the Comp-Available condition were more likely than those in the Comp-Difficult condition to choose according to the broad hypothesis—that is, to predict that the test gardens with red soil were sick and those with brown soil were healthy ($M_{Avail} = 3.41$, $SD_{Avail} = 1.17$; $M_{Diff} = 1.16$, $SD_{Diff} = 1.52$). One-sample $t$-tests compared group performance to a chance level of 2. The Comp-Available group produced the broad pattern of choices at above-chance rates, $t(25) = 6.19$, $p < .001$, $d = 1.21$, while the Comp-Difficult group did so at below-chance rates, $t(24) = -2.77$, $p = .011$, $d = -0.44$. Table 3 shows the

**Table 2** Experiment 1: sample explanations and codes.

| Explanation | Codes |
|---|---|
| Because red soil* | Soil |
| Because the red soil is infecting it* | Soil |
| Red grows sick carrots because of dry dirt; this [*pointing to the sprinkler garden*] because no water; this [*tree*] because it has shadow. Sun will not get in. | Soil, Individual feature |
| It's sick without water because water is making them hydrated. | Individual feature |
| Because this tree is blocking the sunlight to get to the carrots | Individual feature |
| because they didn't get treated that good—water them every single day and give them sunlight | Individual feature |
| Carrots can be sick because they don't get enough water, sunlight; sometimes seeds are not good | Individual feature, Other |
| Because they have doghouse and rocks | Other |
| Because germ makes you sick | Other |
| I don't know | Other |

*Mentioning soil type as the sole explanation.

classification of children's responses. The modal response for the Comp-Available group was to make prediction choices solely based on the common soil type, while that for the Comp-Difficult group was to endorse the context-bound hypothesis.

### Verbal explanation task

We looked first at the relative frequency of mentioning each cause in children's explanations. In the initial explanation (Figure 3, left), the frequency distribution did not significantly differ between conditions, Fisher's exact test, $p = .064$. This was expected since the training procedure to that point was identical for the two groups. In contrast, as predicted, in the final explanation (Figure 3, right), there was a significant condition difference, Fisher's exact test, $p < .001$. Only the Comp-Available group predominantly explained by the soil type. Table 4 gives a sample of children's explanations.

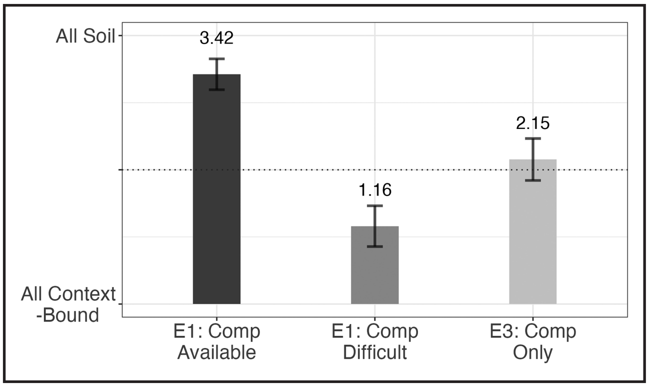


**Figure 2** Results of Experiments 1 and 3: average number of predictions in accord with the broad (soil-based) hypothesis (all soil = 4; all context-bound = 0). Error bars represent standard errors.

**Table 3** Experiments 1 and 3: frequency of the prediction pattern by condition.

| | All soil | All context Bound | Soil + 1 | Sick bias | Healthy bias | Other |
|---|---|---|---|---|---|---|
| E1: Comp-Available | 19 | 2 | 3 | 1 | 0 | 1 |
| E1: Comp-Difficult | 3 | 14 | 3 | 0 | 1 | 4 |
| E3: Comp-Only | 8 | 6 | 1 | 1 | 1 | 9 |

One concern was the higher initial proportion of soil-based explanation for the Comp-Available group than for the Comp-Difficult group (27% vs. 11%). We reran the above analyses among children who did not mention soil type in the initial explanation ($N_{Avail} = 17$; $N_{Diff} = 21$). The same patterns held: the Comp-Available group was more likely than the Comp-Difficult group to favor the soil-type hypothesis in the prediction task ($M_{Avail} = 3.29$, $SD_{Avail} = 1.36$; $M_{Diff} = 0.90$, $SD_{Diff} = 1.34$), $t(34) = 5.43$, $p < .001$, $d = 1.77$, as well as to explain by the soil type in their final explanation.

We also calculated the rate of mentioning soil type as the *sole* cause. In the initial explanation, a Fisher's exact test revealed no difference between the Comp-Available group (9 out of 26) and the Comp-Difficult group (3 out of 25) in their tendency to do so, $p = .10$. In the final explanation, the Comp-Available group (18 out of 26) was significantly more likely than the Comp-Difficult group (4 out of 25) to name soil type as the sole cause of sick carrots, Fisher's exact test, $p < .001$. Thus, the Comp-Available group, for whom comparisons were readily available, showed a marked shift in explanatory preference from having no particular preference to strongly preferring the common soil as the cause of sick carrots. Without such comparison opportunities, the Comp-Difficult group showed no such shift. In sum, whether children had comparisons readily available was a strong predictor of whether they shifted to the broad explanation.

## Experiment 1 discussion

The results of Experiment 1 provide clear evidence for the power of comparison processes in explanation generation. When comparison was readily available, children who explained made predictions according to the broad hypothesis (soil type) and were highly likely to endorse the common soil type in their final verbal explanations. When comparison was difficult, children who explained failed to endorse the broad hypothesis in their predictions and explanations. Thus, having an explanatory goal led to

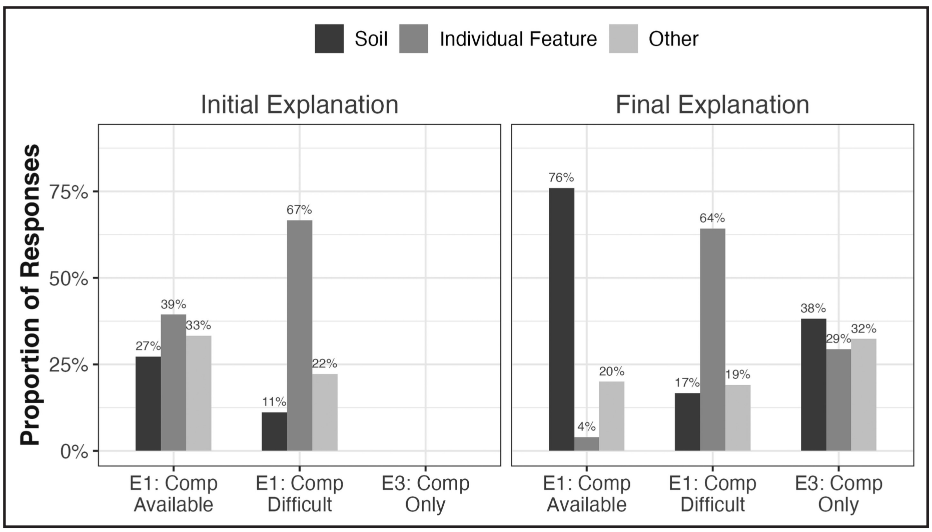


**Figure 3** Experiments 1 and 3: 5- and 6-year olds' rates of mentioning each cause in explanations.

**Table 4** Children's verbal explanations in Experiments 1 and 3.

| Child | Initial explanation | Final explanation |
|---|---|---|
| E1: Comp-Available | | |
| 1 | Bad soil and not enough sun, not enough water | Because red soil* |
| 2 | Soil isn't watered and it needs shades and sun | Because soil is red* |
| 3 | No water | They have different colors (*pointing between red test gardens and brown training gardens*) |
| E1: Comp-Difficult | | |
| 4 | I know why this one is sick. Plants need sun, water and food because it's under the tree. It doesn't have sun | It has litter so it stops the root from getting food. That one has shades, so it doesn't have enough sun |
| 5 | Because it's too hot | (*referring to the garden with tree*) stuff don't grow in shade; (*the garden with rocks*) bad bacteria were in the sand; (*the garden with sprinkler*) broken sprinkler probably making the sand bad |
| 6 | They need water | It doesn't have water; need to fix the hose |
| E3: Comp-Only | | |
| 7 | N/A | This one has a broken sprinkler; the other one has red soil |
| 8 | N/A | because this tree looks like it's falling down; this broken sprinkler the carrots just don't want to grow |
| 9 | N/A | (*referring to the garden with rocks*) maybe I don't know; (*the garden with sprinkler*) definitely because there is broken sprinkler. Need to grow plant |

*Mentioning soil type as the sole explanation.

the broad hypothesis *only* when children could readily compare the items—suggesting that comparison was critical in allowing children to arrive at a broad explanation.

A possible reservation concerning the findings is that our manipulation of whether the gardens were shown simultaneously might have created an extra memory load for the Comp-Difficult group, thus impeding their capacity to understand the task. We think this is unlikely because both groups showed excellent memory performance; they remembered the individual training gardens and whether they had sick or healthy carrots.

We are suggesting that in our Experiment 1 (and in Walker et al.'s (2017) study), when 5- and 6-year olds were asked to explain the pattern, they spontaneously drew comparisons when such comparisons were readily available and that this enabled them to perceive commonalities that supported their shift to the broad explanation. These results link with prior findings (Edwards et al, 2019; Hoyos & Gentner, 2017) in providing support for the claim that comparison is integral to the power of explanation in revealing broad explanatory patterns.

In Experiment 2, we extend our logic to younger children. We ask whether younger children will arrive at the broad hypothesis if encouraged to make comparisons. As reviewed above, the 4-year olds in Walker et al.'s (2017) study failed to show a preference for broad explanations when given the same comparison-available condition that led 5- and 6-year olds to adopt the broad hypothesis. The authors speculated that the 4-year olds might have lacked sufficient domain knowledge about plant growth. We agree and suggest further that this lack of knowledge might have prevented 4-year olds from recognizing the comparison opportunity despite its availability. This is consistent with findings that domain knowledge plays a key role in driving spontaneous comparison (Gentner & Hoyos, 2017; Gentner & Rattermann, 1991; Goldwater & Gentner, 2015; Paik & Mix, 2006; Rottman et al., 2012). There is also considerable research showing that comparison can be explicitly encouraged to benefit learning, both in adults (Edwards et al., 2019; Gentner et al., 2009; Gick & Holyoak, 1983; Goldwater & Gentner, 2015; Kurtz et al., 2013; Kurtz et al., 2001; Loewenstein et al., 2003) and children (Alfieri et al., 2013; Begolli & Richland, 2016; Gentner et al., 2011; Gentner & Namy, 1999; Honomichl & Chen, 2006; Mason, 2004; Mix et al., 2018; Murphy et al., 2021; Rittle-Johnson & Star, 2009).

In Experiment 2, we prompted 4-year olds to engage in comparison and asked whether this increased their likelihood of identifying the common soil type pattern. We did this both by directly inviting children to compare the healthy and the sick gardens and also by creating perceptual conditions favorable to comparison: specifically, by showing the training gardens simultaneously and by aligning them spatially, such that gardens of the same soil type were juxtaposed (Matlen et al., 2020; Zheng et al., 2022). The control group completed the same procedure as the Comp-Available condition from E1.

# Experiment 2

## Method

### Participants

Sixty children participated in the study ($M = 4;6$ years, range = 4;00 to 5;00 years; 28 girls; 53% White, 25% Asian, 8% Black, 7% Hispanic, 7% Mixed; data collected 2024): 29 in the Explanation + Comparison Available condition [Comp-Available] condition and 31 in the Explanation + Comparison Prompted condition [Comp-Prompted]. Families were recruited for a Zoom study through Instagram advertisements and compensated with gift cards. The participation criteria were set to be based in the United States with English as the primary language. We did not collect information about families' socioeconomic status. Two children in the Comp-Prompted condition and one in the Comp-Available condition were excluded due to memory failure (more than one error on the memory task), resulting in a final $N_{Avail} = 28$ and $N_{Prompt} = 29$.

### Materials and procedure

The materials from E1 were digitized as PowerPoint slides for Zoom. The procedure for the Comp-Available condition was identical to that for the same condition in E1, except for minor adaptations noted below. The Comp-Prompted differed from the Comp-Available condition in the Training phase, where the critical manipulation of the comparison prompt occurred. The Memory and Test phases were the same between conditions.

*Training*

The conditions differed in how they received the training gardens. As in E1, the Comp-Available children saw the four gardens sequentially, each on a different slide (alternating between healthy and sick gardens). For each, the experimenter said, for example, "In this garden, there is an X, and the soil is Y," and then the picture of a healthy or sick carrot was revealed. For technical reasons, children did not select which leaf clump to examine. Instead, they were guided to check a prespecified clump. The healthy or sick carrot picture appeared upon click. Thus, children learned that the two brown soil gardens grew healthy carrots, and the two red soil gardens grew sick gardens. At the last garden, the experimenter asked the child for their initial explanation, "Why do you think some carrots are sick?"

The Comp-Prompted children saw all four gardens together arranged as quadrants of a bigger rectangle (Figure 1, left). In addition, the experimenter first described the two brown soil gardens and then the two red soil ones. Both procedures were the same as in Walker et al. (2017) and different from those in the Comp-Available condition in E1 and E2 (Table 1). Importantly, after learning about each garden, children received an explicit prompt to compare. The experimenter said, "Let's put all the healthy ones together!", and the two brown soil gardens then moved to the left side of the screen. The experimenter then asked, "What's alike about these two gardens?" Any response from the child was accepted, as the goal was to encourage comparison. The same steps were then repeated for the two sick gardens, which were on the right side of the screen. Last, the experimenter asked for the child's initial explanation (with all four gardens present, just as in Walker et al.).

*Memory*

For both conditions, the Memory and Test phases were the same as in the Comp-Available condition in E1. In the Memory task, the training gardens were shown one after the other. For each, the experimenter asked, "Do you remember this garden? Does it grow healthy carrots or sick carrots?" Wrong answers were corrected. Then the garden was moved to the side and remained visible until the end of the study. The gardens were shown in a fixed order, alternating between healthy and sick ones. At the end of the Memory phase, the four gardens formed a column on the side.

*Prediction task*

With the training gardens visible on the side, children were shown the test gardens one at a time, with the order randomized. For each garden, the child was asked whether it would grow healthy or sick carrots. Afterward, children were asked for their final explanation with all four test gardens present. The experimenter circled the gardens children had predicted to grow sick carrots and asked, "Why do you think these grow sick carrots?"

## Experiment 2 results

### Prediction task

(Figure 4) A $t$-test revealed a significant effect of condition, $t(55) = 4.08$, $p < .001$, $d = 1.08$. Children in the Comp-Prompted condition were more likely than those in the Comp-Available condition to choose according to the broad (soil-type) hypothesis ($M_{Prompt} = 3.00$, $SD_{Prompt} = 1.54$; $M_{Avail} = 1.29$, $SD_{Avail} = 1.63$). Further, one-sample $t$-tests showed that the Comp-Prompted group produced this broad pattern of choices at above-chance rates, $t(28) = 3.51$, $p = .002$, $d = 0.65$, while the Comp-Available group did so at below-chance rates, $t(27) = -2.32$, $p = .028$, $d = -0.44$. Table 5 shows the classification of children's responses. The modal response for the Comp-Prompted group was to choose based on the broad hypothesis (i.e., the soil type), while that for the Comp-Available group was to choose according to the context-bound hypothesis of individual features. Thus, although all children were asked to explain the pattern, only those who were encouraged to engage in comparison arrived at the broad hypothesis.

### Verbal explanation task

We looked first at the relative frequency of mentioning each cause in children's explanations. Fisher's exact tests showed that the two groups differed in their initial explanation (Figure 5, left), $p = .044$. However, neither group mentioned soil type as the primary cause: the modal response of the Comp-Available group was to mention irrelevant causes, and that of the Comp-Prompted group was to mention lack of water or sunlight. The trend of a higher proportion of soil-based explanations in the Comp-Prompted group might have resulted from the fact that the Comp-Prompted group had already been invited to compare across the gardens before being asked for the initial explanation. We reran the Prediction task analysis among children who did not mention soil type in the initial explanation ($N_{Prompt} = 17$; $N_{Avail} = 24$). The same pattern held: the Comp-Prompted group was more likely to favor the soil-type hypothesis than was the Comp-Available group, $t(39) = 2.59$, $p = .014$, $d = 0.82$. However, contrary to our hypothesis, the two conditions did not differ in the final explanation (Figure 5, right), $p = .33$.

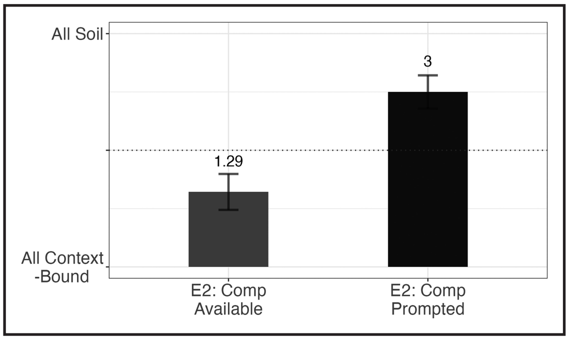


**Figure 4** Results of Experiment 2: average number of predictions in accord with the broad hypothesis (all soil = 4; all context-bound = 0). Error bars represent standard errors.

We also calculated the rate of mentioning soil type as the *sole* cause. There was no difference between conditions in either the initial explanation (Comp-Available: 3 out of 28; Comp-Prompted: 7 out of 29), Fisher's Eeact test, $p = .3$, or the final explanation (Comp-Available: 8 out of 28; Comp-Prompted: 14 out of 29), Chi-squared test, $p = .21$.

Thus, the planned analysis for verbal response failed to provide support for the claim that prompting comparison matters for explanation. However, as indicated in Figure 5 and Table 6, the high proportion of "Other" responses suggests that some 4-year-old children might have found the explanation task challenging. We further explored this issue by dividing children based on their performance in the Prediction task. The pattern was telling. Those who made "All Soil" predictions ($N_{Avail} = 6$; $N_{Prompt} = 18$) predominantly mentioned soil type in the final explanation (Comp-Available: 83% *Soil*, 17% *Other*; Comp-Prompted: *71*% Soil, 10% *Feature,* 19% *Other*). This is a marked gain over these children's initial explanations, in which 0% and 39% of the responses were soil type-based for the Comp-Available and Comp-Prompted groups, respectively. In contrast, for children who did not make "All Soil" predictions ($N_{Avail} = 22$; $N_{Prompt} = 11$), their final explanations were far less likely to mention soil-type (Comp-Available: 16% *Soil,* 38% *Feature,* 46% *Other*; Comp-Prompted: 8% *Soil*, 33% *Feature, 58*% Other). Indeed, half of their responses focused on irrelevant causes. In sum, children's verbal responses echoed their prediction performance: those who showed a clear preference for the common soil-type hypothesis in making predictions also mentioned soil-type more in the final explanation. This was the case for the Comp-Available group as well as the Comp-Prompted group (who were three times more likely than the former group to make "All Soil" predictions)

**Table 5** Experiment 2: frequency of the prediction pattern by condition.

| | All soil | All context-Bound | Soil + 1 | Sick bias | Healthy bias | Other |
|---|---|---|---|---|---|---|
| E2: Comp-Available | 6 | 15 | 0 | 0 | 2 | 5 |
| E2: Comp-Prompted | 18 | 5 | 3 | 1 | 1 | 1 |

## Experiment 2 discussion

The results of Experiment 2 provide further support for our main hypothesis: that comparison processes are critical to explanation's effect in revealing broad general patterns. In our Comp-Available condition, in which comparison was possible but not encouraged, 4-year olds failed to arrive at the broad (soil-based) hypothesis, replicating Walker et al.'s (2017) findings for this age group. But when comparison was encouraged, 4-year-olds in our Comp-Prompted condition did form the broad hypothesis, as reflected in their performance on the prediction task. The difference between the two groups is striking: the Comp-Prompted group made over twice as many predictions consistent with the broad hypothesis as did the Comp-Available group.

To put these results in perspective, recall that in Walker et al.'s study (using what in our terms was a "Comparison Available" procedure), 5–6-year olds formed the broad hypothesis and 4-year olds did not. In our Experiment 2, we found that increasing the likelihood of comparison (as in Explanation + Comparison Prompted) led to an increase in the proportion of 4-year old children who formed the broad hypothesis (relative to Walker et al.'s results, as well as to the results of our control condition, Explanation + Comparison Available). These results dovetail with those of Experiment 1, in which we found that decreasing the availability of comparison (Explanation-Only) led to a decrease in the proportion of 5–6-year-old children who formed the broad hypothesis (relative to Walker et al.'s results and to our replication of those results, Explanation + Comparison Available). Overall, these findings reveal an interesting developmental pattern and provide strong evidence for the role of comparison in allowing children to achieve broad causal insight in explanations.

In Experiment 3, we pose a further question. We have proposed that the goal of explaining a phenomenon actively engages a search for informative comparisons. However, it is worth considering a more extreme possibility. Could it be that *all* the benefits of explanation are due to comparison? This would mean that, as long as the comparison opportunity is present,

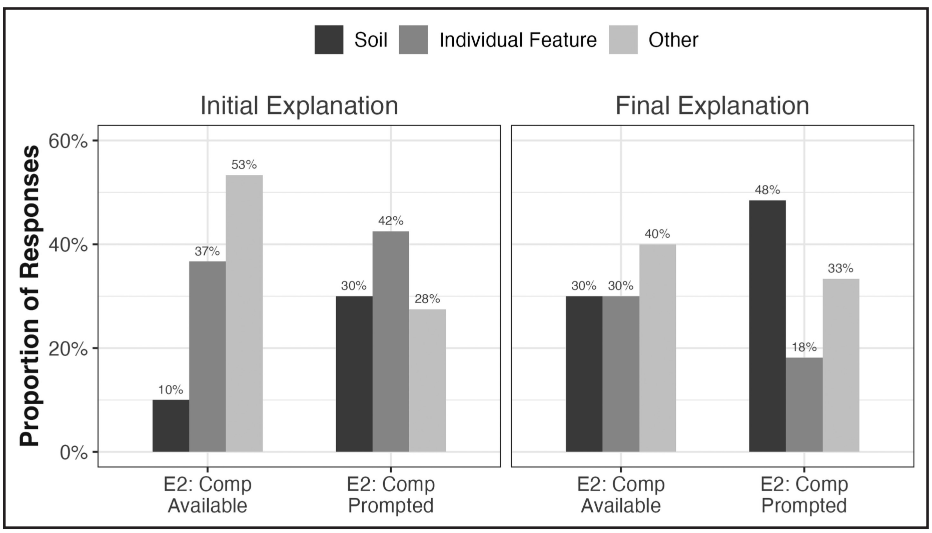


**Figure 5** Experiment 2: 4-year-olds' rates of mentioning each kind of cause in explanations.

**Table 6** Children's explanations in Experiment 2.

| Child | Initial explanation | Final explanation |
|---|---|---|
| Comp-Prompted | | |
| 1 | Because the gardens are not doing very well | Because they are not doing well |
| 2 | Because they are inside the dirt too long | Because the dirt is not red right not; long time ago dirt is red, and they will grow sick carrots |
| 3 | If sun makes them so hot, too much shade, or not much water | Because they don't go sick gardens |
| Comp-Available | | |
| 4 | Because it's dirty it has dirty soil | Because they have dirty nature and dirty trash and way too gross |
| 5 | Because they are not healthy; they are sick | Because it's not healthy we can't eat them |
| 6 | Because they get bad food on the ground it makes it dirty, and you can't eat it | Because they grew on the ground |

children would not need explanation to endorse the broad pattern. Experiment 3 examines whether comparison by itself is enough to drive children's preference for broad regularities. A new group of 5- and 6-year-olds, the Comparison-Only condition, completed the study in the same way as did the Comp-Available condition in Experiment 1, except for one critical difference: this group did not explain during training why some carrots were sick.

# Experiment 3

## Method

### Participants

Twenty-six 5- and 6-year-old children participated, all in the Comparison-Only [Comp-Only; labeled in contrast to the Comp-Available condition from E1] condition ($M = 6;00$ years, range $= 5;1$ to 7;00 years; 13 girls; 61% White, 12% Asian, 27% Mixed; data collected 2024). No one was excluded. Children were recruited using the same methods as in Experiment 1.

### Procedure

Children received the identical procedure as did the Comp-Available condition from E1, except in one important respect: they were not asked for the initial explanation during training. Children were asked to help Mr. Farmer figure out what made the carrots healthy or sick. They saw each of the training gardens one by one (in a fixed sequence) and learned that the two brown soil gardens (with a doghouse and rocks) grew healthy carrots while the two red soil gardens (with a broken sprinkler and a shady tree) grew sick carrots. After the last garden, the children heard, "Let's see how well you remember the gardens," instead of being asked to explain why the carrots were sick, as in E1. The rest of the procedure was identical to that of the Comp-Available condition in E1.

*Memory*

The training gardens were shown one after the other, and the child was asked to remember the state of the carrots for each training garden. Children's memory was good: only three children each made a single error. Each garden was then pushed to the side and remained in view; thus, comparisons were readily available during the Prediction task.

*Prediction task*

As in Experiments 1 and 2, the four test gardens reversed the association between soil type and specific features. For each of the test gardens, the child was asked whether it would grow healthy or sick carrots. Last, each child was shown the test garden(s) that they had predicted would grow sick carrots and asked, "Why do you think they grow sick carrots?" This constituted their final (and only) explanation.

# Experiment 3 results

### Prediction task

Figure 2 shows children's performance (labeled as Comp-Only condition) ($M_{C\text{-}Only} = 2.15$, $SD_{C\text{-}Only} = 1.59$) along with the two conditions from Experiment 1. For the prediction task, we again first carried out an omnibus ANOVA including condition (Comp-Only, Comp-Available, or Comp-Difficult), age group (5- or 6-year olds), and their interaction as independent variables. The results showed that neither age, $F(1, 71) = 0.31$, $p = .58$, nor its interaction with condition, $F(2, 71) = 2.08$, $p = .13$, was significant. Importantly, there was an effect of condition, $F(2, 71) = 16.19$, $p < .001$. A follow-up $t$-test revealed that the Comp-Only group, who did not engage in explanation, was significantly less likely to follow the broad soil-type hypothesis in their predictions than was the Comp-Available group from E1, $t(45) = -3.27$, $p = .002$, $d = -0.91$. In addition, a one-sample $t$-test showed that the Comp-Only group was at chance in predicting that the test gardens with red soil were sick, $p = .63$. Table 3 shows the classification of children's responses. Children in the Comp-Only condition gave roughly the same number of broad and context-bound predictions, suggesting that this group did not form a strong bias to reason based on the common soil type.

### Verbal explanation task

We compared the performance of children in Experiment 3 on the final explanation task with that of the two groups in Experiment 1 (Figure 3). The Fisher's exact test showed that the Comp-Only group and the Comp-Available group differed in the relative frequency of mentioning each cause, $p = .007$. Only 10 out of 26 children in the Comp-Only group (Experiment 3) reported soil type as the *sole* cause for what made carrots sick, as compared to 18 out of 26 in the Comp-Available group from E1, though this difference was not significant, $\chi^2 = 3.79$, $p = .052$.

**Table 7** Children's tendency to support the broad explanation when asked to explain, across age and comparison opportunity.

| | | Comparison difficult | Comparison available | Comparison prompted |
|---|---|---|---|---|
| 5- and 6-year olds | Walker et al. (2017) | | ✓ | |
| | Present E1 | X | ✓ | |
| 4-year olds | Walker et al. (2017) | | X | |
| | Present E2 | | X | ✓ |

Checkmarks denote success at arriving at the broad explanation.

We also explored the differences between the Comp-Only condition and the Comp-Difficult condition. One non-predicted result was that the Comp-Only group was more likely to follow the broad hypothesis in the Prediction task than was the Comp-Difficult group, $t(49)=2.28$, $p=.027$, $d=0.64$. Likewise, the Comp-Only group was more likely to mention soil type in their explanations than was the Comp-Difficult, Fisher's exact test, $p=.009$. However, the two groups did not differ in reporting soil type as the *sole* cause, Fisher's exact test, $p=.12$. As suggested by Figure 3, the Comp-Only group did not show a strong preference for any particular cause in their verbal explanations, while the Comp-Difficult group tended to explain by the features of lack of sunlight and lack of water.

## Experiment 3 discussion

The results of Experiment 3 showed that without a prompt to explain, 5–6-year-old children did not focus on the broad cause of soil type in their prediction and final explanation, despite having readily available comparisons. These findings argue against the extreme possibility raised by Experiments 1 and 2, that the availability of comparisons is sufficient to allow children to arrive at the broad explanation. Children must also have the desire to understand the phenomena. Asking children to explain, as in our Experiments 1 and 2 and in Walker et al.'s study, provides a motivation to understand and therefore to engage in comparison.

Experiment 3 tested the extreme possibility that comparison by itself would be sufficient to support children's arriving at the broad explanation. We gave a new group of 5- and 6-year olds the same opportunity to compare gardens as in our Experiment 1 (and Walker et al., 2017) but did not ask them to explain. These children performed at chance in the prediction task; they were significantly less likely to adopt the broad (soil-type) hypothesis than were children who had an explanatory goal as well as available comparisons. Likewise, when later asked to explain, they showed little evidence of having formed a preference for the broad hypothesis. We propose that having an explanatory goal plus an available set of relevant comparisons may be the ideal state for achieving a broad hypothesis. Without the goal of explaining, children may fail to utilize available comparison opportunities (Sidney et al., 2015). Similarly, when comparison is unavailable, children who are trying to explain something may resort to salient object features (e.g., a broken sprinkler; Cimpian, 2015).

These results dovetail with prior studies that have found a connection between self-explanation and comparison (Chin-Parker & Bradner, 2017; Edwards et al., 2019; Hilton & Slugoski, 1986; Hoyos & Gentner, 2017; Landy & Hummel, 2010; Sidney et al., 2015). For example, Hummel et al. (2014) note that both explanation and analogy are inductive and knowledge-intensive processes and suggest that models of analogical processing can be extended to describe explanatory processes.

# General discussion

In three experiments, we found evidence for our two claims (1) that generating explanations (often) recruits comparison and (2) that the comparison process acts to improve explanation-based learning by revealing broad common patterns. In Experiments 1 and 2, we replicated Walker et al.'s (2017) results by showing that when asked to explain (with available comparisons), 5- and 6-year-old children reasoned by the common soil type pattern, whereas 4-year-old children did not. We then assessed the role of comparison in two ways: by reducing comparison likelihood for the older group and by increasing it for the younger group. When comparison likelihood was reduced (Experiment 1), 5- and 6-year olds no longer endorsed the broad hypothesis. When comparison likelihood was increased (Experiment 2), even 4-year olds arrived at the broad causal insight in making predictions. Together, these findings provide strong support for our claim that comparison is critical to explanation's success in leading to broad hypotheses. Across our studies and those of Walker et al. (2017), we see a consistent pattern (Table 7). Given an explanatory goal, older children are more likely to arrive at the broad hypothesis than younger children. But for both age groups, the availability of comparisons matters in whether children reach this hypothesis.

## Further issues

One open question concerns the nature of the connection between explanation and comparison processing. It may be that children—especially older children—view comparison as a useful learning strategy and actively deploy it when seeking to explain something—for example, by comparing positive instances (e.g., the two healthy gardens) vs. negative instances (the two sick gardens). This would fit with the results of the Edwards et al. (2019) study, in which adults who were asked to explain a category rule reported engaging in comparison to a greater extent than those who were explicitly told to compare items. However, the connection between explanation and comparison might be less deliberate. Perhaps the goal to explain arouses children's curiosity, leading them to attend more to the scene and to notice and process comparisons. Consistent with this possibility, Baillargeon and colleagues have suggested that even infants search for explanations when encountering counterintuitive physical events (Baillargeon & DeJong, 2017; Wang & Baillargeon, 2008). We conjecture that both these paths occur, at least for older children and adults.

Our studies show that having an explanatory goal can lead to invoking comparison processes. However, we suggest that the opposite path can also occur—comparison itself can excite curiosity and lead the child to investigate further. Consistent with this idea, we found an interesting pattern in Experiment 3. As noted earlier, 5–6-year olds who had available comparisons but were not asked to explain (Experiment 3) were significantly less likely to form the broad hypothesis than were the same-age children who were given an explanatory goal as well as available comparisons (Experiment 1)—evidence that the goal to explain matters. However, there was a further intriguing finding: children in Experiment 3 were significantly *more* likely to favor the broad hypothesis than children in the Explanation-only group in Experiment 1. In other words, children were more likely to form the broad hypothesis when given comparison by itself than when given an explanation task by itself. We speculate that spontaneous comparison led children to notice commonalities, making them curious to understand the whole pattern. Prior studies have identified a number of factors that contribute to the likelihood of spontaneous comparison, including high perceptual similarity (Gentner et al., 2017; Gentner & Hoyos, 2017; Gentner & Rattermann, 1991), spatiotemporal juxtaposition (Begolli & Richland, 2016; Christie & Gentner, 2010; Rittle-Johnson & Star, 2009), and spatial alignment (Matlen et al., 2020; Zheng et al., 2022). In addition, using the same label for two things provides a strong incentive for children to compare them (Christie & Gentner, 2010; Gentner & Namy, 1999; Zheng et al., 2025). In the current study, labeling both red soil gardens as being sick might have motivated comparison to some extent, though the results show that this was not sufficient to shift children's preference.

Finally, although we have focused here on the role of comparison in revealing common relational structures, the comparison process can also contribute to explanation by rendering important *differences* more salient (Chin-Parker & Bradner, 2017). For example, Hoyos and Gentner (2017) tested whether explanation and comparison can benefit learning of differences. They showed 6-year-old children a pair of model buildings—one with a diagonal brace (which was stable) and one with a horizontal crosspiece (which was unstable). Children were asked to explain why the stable building was stronger. For half the children, the buildings looked similar and were easy to align. For the other half, the two buildings were difficult to align. The results showed that ease of alignment was a strong predictor of identifying the correct relational difference: that the building with the diagonal brace was stronger. Indeed, children who saw the hard-to-align pair mostly explained by overall shape and surface differences. The authors suggested that when given two highly similar buildings, children readily engaged in structural alignment, thus highlighting both common relational structure and alignable differences—that is, differences that play corresponding roles in the aligned structure (Markman & Gentner, 1993; Sagi et al., 2012)—in this case, the difference between the diagonal brace and the horizontal piece. In sum, both contrastive comparison and commonality-focused comparison can contribute to explanation generation.

## Implications for education

A wide range of research has supported the idea that comparison-based instructions can be highly effective in promoting abstract knowledge acquisition (Alfieri et al., 2013; Begolli & Richland, 2016; Gadgil et al., 2012; Goldwater & Gentner, 2015; Honomichl & Chen, 2006; Jee & Anggoro, 2019; Mix et al., 2018; Rittle-Johnson & Star, 2009). The benefits of comparison extend to both classroom learning and everyday contexts. For example, Goldwater and Gentner (2015) used comparison to train college students to better understand the causal structures of diverse everyday phenomena (e.g., a positive feedback loop in global warming and economics). At baseline, students predominantly sorted events in terms of concrete domain, rather than by the underlying causal structures. Even with direct explications of how examples fit the causal schemas, students mostly failed to transfer this pattern to new cases. In contrast, analogical comparison between parallel examples increased students' propensity to detect the common causal structures and allowed them to transfer this insight across diverse events. This speaks to the power of even a brief comparison training experience.

To inform effective teaching, it is important to empirically establish how students can carry out comparisons fluently. Through a meta-analysis, Alfieri et al. (2013) confirmed that case comparisons generally led to improved learning outcomes over other forms of classroom instruction. They further identified four variables that reliably moderated this effect: instructional comparisons are most effective when (1) students are instructed to search for commonalities, (2) abstract principles are provided after the comparisons, (3) there is perceptual content to align, and (4) testing is immediate.

An equally important question is whether and how learners come to spontaneously recognize available comparison opportunities. Indeed, recent research suggests that college students on average miss half of comparisons in expository science texts, even when being explicitly told to seek out comparisons (Mo et al., in preparation). The current work suggests that imposing a goal of explaining can actively motivate learners to seek relevant comparisons, thereby increasing the likelihood of noticing and learning from them. This idea dovetails with previous research that links everyday elaborative *wh*-questions to a variety of developmental benefits (Fivush et al., 2006; Willard et al., 2019). It is possible that through elaborative conversations, the habitual practice of connecting self-explanation with comparison becomes a habitual mindset and confers long-term cognitive and educational benefits. Future research can look at the longitudinal relationship between self-explanation and comparison.

# Conclusion

Explanation generation is a productive learning process in both children and adults. Our results suggest that its power stems in part from invoking comparison processes. Engaging in explanation can instill curiosity and motivate children to try to understand phenomena. To achieve this, explainers often seek available comparisons in the environment or in their own prior knowledge. Through the structural alignment process of comparison, learners come to notice common relational patterns that provide the material for broad, satisfying explanations. In sum, we suggest that explanation and comparison are mutually scaffolding learning processes.

# Supplementary material

Supplementary material is available at *Child Development* online

## Data availability

The study is preregistered. The preregistration is available at https://doi.org/10.17605/OSF.IO/EW6HJ

The data reported in this article, the analytic code necessary to reproduce the analyses, and the materials are publicly accessible. They are available through the Open Science Framework at https://osf.io/5zd36/?view_only=f5d34b4ca996475eb5671add954c5a57

## Author contributions

Yinyuan Zheng (Conceptualization, Methodology, Validation, Writing—original draft, Writing—review & editing [equal], Data curation, Formal analysis, Investigation, Project administration, Visualization [lead]), and Dedre Gentner (Conceptualization, Methodology, Validation, Writing—original draft, Writing—review & editing [equal], Data curation, Formal analysis, Investigation, Project administration, Visualization [supporting], Funding acquisition, Resources, Supervision [lead])

## Funding

This project was funded by the Spatial Intelligence and Learning Center (SILC), an National Science Foundation grant, SBE-0541957, to the second author. National Science Foundation had no involvement in the study design; in the collection, analysis and interpretation of data; in the writing of the report; or in the decision to submit the article for publication.

## Conflicts of interest

None declared.

## Acknowledgments

We thank David Uttal and Sid Horton and members of Northwestern's Cognition and Language laboratory for helpful comments and suggestions.

## References

Alfieri, L., Nokes-Malach, T. J., & Schunn, C. D. (2013). Learning through case comparisons: A meta-analytic review. *Educational Psychologist*, *48*, 87–113. https://doi.org/10.1080/00461520.2013.775712

Baillargeon, R., & DeJong, G. F. (2017). Explanation-based learning in infancy. *Psychonomic Bulletin & Review*, *24*, 1511–1526. https://doi.org/10.3758/s13423-017-1334-4

Begolli, K. N., & Richland, L. E. (2016). Teaching mathematics by comparison: Analog visibility as a double-edged sword. *Journal of Educational Psychology*, *108*, 194. https://doi.org/10.1037/edu0000056

Bisra, K., Liu, Q., Nesbit, J. C., Salimi, F., & Winne, P. H. (2018). Inducing selfexplanation: A meta-analysis. *Educational Psychology Review*, *30*, 703–725. https://doi.org/10.1007/s10648-018-9434-x

Brockbank, E., Lombrozo, T., Gopnik, A., & Walker, C. M. (2023). Ask me why, don't tell me why: Asking children for explanations facilitates relational thinking. *Developmental Science*, *26*, Article e13274. https://doi.org/10.1111/desc.13274

Busch, J. T., Willard, A. K., & Legare, C. H. (2018). Explanation scaffolds causal learning and problem solving in childhood. In M. Saylor, & P. Ganea (Eds.), *Active learning from infancy to childhood: Social motivation, cognition, and linguistic mechanisms* (pp. 113–127). Springer International Publishing.

Callanan, M. A., & Oakes, L. M. (1992). Preschoolers' questions and parents' explanations: Causal thinking in everyday activity. *Cognitive Development*, *7*, 213–233. https://doi.org/10.1016/0885-2014(92)90012-g

Chen, Z. (1999). Schema induction in children's analogical problem solving. *Journal of Educational Psychology*, *91*, 703–715. https://doi.org/10.1037/0022-0663.91.4.703

Chi, M. T. H. (2000). Self-explaining: The dual processes of generating inference and repairing mental models. In R. Glaser (Ed.), *Advances in instructional psychology: Educational design and cognitive science* (Vol. *5*, pp. 161–238). Lawrence Earlbaum.

Chi, M. T. H., Bassok, M., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science*, *13*, 145–182. https://doi.org/10.1016/0364-0213(89)90002-5

Chi, M. T. H., De Leeuw, N., Chiu, M. H., & LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science*, *18*, 439–477. https://doi.org/10.1016/0364-0213(94)90016-7

Chin-Parker, S., & Bradner, A. (2017). A contrastive account of explanation generation. *Psychonomic Bulletin & Review*, *24*, 1387–1397. https://doi.org/10.3758/s13423-017-1349-x

Christie, S., & Gentner, D. (2007). Relational similarity in identity relation: The role of language. In Vosniadou, S., & Kayser, D. (Eds.), *Proceedings of the Second Europea Cognitive Science Conference* (pp. 401–406). Lawrence Erlbaum Associates.

Christie, S., & Gentner, D. (2010). Where hypotheses come from: Learning new relations by structural alignment. *Journal of Cognition and Development*, *11*, 356–373. https://doi.org/10.1080/15248371003700015

Cimpian, A. (2015). The inherence heuristic: Generating everyday explanations. In R. A. Scott, & S. M. Kosslyn (Eds.), *Emerging trends in the social and behavioral sciences* (pp. 1–15). Wiley.

Dixon, J. A., & Bangert, A. S. (2004). On the spontaneous discovery of a mathematical relation during problem solving. *Cognitive Science*, *28*, 433–449. https://doi.org/10.1207/s15516709cog2803_6

Edwards, B. J., Williams, J. J., Gentner, D., & Lombrozo, T. (2019). Explanation recruits comparison in a category-learning task. *Cognition*, *185*, 21–38. https://doi.org/10.1016/j.cognition.2018.12.011

Fivush, R., Haden, C. A., & Reese, E. (2006). Elaborating on elaborations: Role of maternal reminiscing style in cognitive and socioemotional development. *Child Development*, *77*, 1568–1588. https://doi.org/10.1111/j.1467-8624.2006.00960.x

Forbus, K. D., Ferguson, R. W., Lovett, A., & Gentner, D. (2017). Extending SME to handle large-scale cognitive modeling. *Cognitive Science*, *41*, 1152–1201. https://doi.org/10.1111/cogs.12377

Gadgil, S., Nokes-Malach, T. J., & Chi, M. T. H. (2012). Effectiveness of holistic mental model confrontation in driving conceptual change. *Learning and Instruction*, *22*, 47–61. https://doi.org/10.1016/j.learninstruc.2011.06.002

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science*, *7*, 155–170. https://doi.org/10.1207/s15516709cog0702_3

Gentner, D. (2003). Why we're so smart. In D. Gentner, & S. Goldin-Meadow (Eds.), *Language in mind: Advances in the study of language and thought* (pp. 195–236). MIT Press https://doi.org/10.7551/mitpress/4117.003.0015

Gentner, D. (2010). Bootstrapping the mind: Analogical processes and symbol systems. *Cognitive Science*, *34*, 752–775. https://doi.org/10.1111/j.1551-6709.2010.01114.x

Gentner, D., Anggoro, F., & Klibanoff, R. S. (2011). Structure mapping and relational language support children's learning of relational categories. *Child Development*, *82*, 1173–1188. https://doi.org/10.1111/j.1467-8624.2011.01599.x

Gentner, D., & Forbus, K. (2025). The structure-mapping engine: A multidecade interaction between psychology and artificial intelligence. *Current Directions in Psychological Science*, 1–8. https://doi.org/10.1177/09637214251395678.

Gentner, D., & Hoyos, C. (2017). Analogy and abstraction. *Topics in Cognitive Science*, *9*, 672–693. https://doi.org/10.1111/tops.12278

Gentner, D., Loewenstein, J., Thompson, L., & Forbus, K. D. (2009). Reviving inert knowledge: Analogical abstraction supports relational retrieval of past events. *Cognitive Science*, *33*, 1343–1382. https://doi.org/10.1111/j.1551-6709.2009.01070.x

Gentner, D., & Namy, L. L. (1999). Comparison in the development of categories. *Cognitive Development*, *14*, 487–513. https://doi.org/10.1016/s0885-2014(99)00016-7

Gentner, D., & Rattermann, M. J. (1991). Language and the career of similarity. In S. A. Gelman, & J. P. Byrnes (Eds.), *Perspectives on thought and language: Interrelations in development* (pp. 225–277). Cambridge University Press eBooks.

Gentner, D., & Toupin, C. (1986). Systematicity and surface similarity in the development of analogy. *Cognitive Science*, *10*, 277–300. https://doi.org/10.1016/s0364-0213(86)80019-2

Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology*, *15*, 1–38. https://doi.org/10.1016/0010-0285(83)90002-6

Goldwater, M. B., & Gentner, D. (2015). On the acquisition of abstract knowledge: Structural alignment and explication in learning causal system categories. *Cognition*, *137*, 137–153. https://doi.org/10.1016/j.cognition.2014.12.001

Hilton, D. J., & Slugoski, B. R. (1986). Knowledge-based causal attribution: The abnormal conditions focus model. *Psychological Review*, *93*, 75–88. https://doi.org/10.1037/0033-295x.93.1.75

Honomichl, R. D., & Chen, Z. (2006). Learning to align relations: The effects of feedback and self-explanation. *Journal of Cognition and Development*, *7*, 527–550. https://doi.org/10.1207/s15327647jcd0704_5

Hoyos, C., & Gentner, D. (2017). Generating explanations via analogical comparison. *Psychonomic Bulletin & Review*, *24*, 1364–1374. https://doi.org/10.3758/s13423-017-1289-5

Hummel, J. E., Licato, J., & Bringsjord, S. (2014). Analogy, explanation, and proof. *Frontiers in Human Neuroscience*, *8*, 867. https://doi.org/10.3389/fnhum.2014.00867

Jee, B. D., & Anggoro, F. (2019). Relational scaffolding enhances children's understanding of scientific models. *Psychological Science*, *30*, 1287–1302. https://doi.org/10.1177/0956797619864601

Keil, F. C. (2006). Explanation and understanding. *Annual Review of Psychology*, *57*, 227–254. https://doi.org/10.1146/annurev.psych.57.102904.190100

Kroupin, I., & Carey, S. (2021). Population differences in performance on Relational Match to Sample (RMTS) sometimes reflect differences in inductive biases alone. *Current Opinion in Behavioral Sciences*, *37*, 75–83. https://doi.org/10.1016/j.cobeha.2020.11.006

Kurtz, K. J., Boukrina, O., & Gentner, D. (2013). Comparison promotes learning and transfer of relational categories. *Journal of Experimental Psychology*, *39*, 1303–1310. https://doi.org/10.1037/a0031847

Kurtz, K. J., Miao, C., & Gentner, D. (2001). Learning by analogical bootstrapping. *Journal of the Learning Sciences*, *10*, 417–446. https://doi.org/10.1207/s15327809jls1004new_2

Landy, D., & Hummel, J. E. (2010). Explanatory reasoning for inductive confidence. In S. Ohlsson, & R. Catrambone (Eds.), *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (pp. 2894–2899). Cognitive Science Society.

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

Legare, C. H., Sobel, D. M., & Callanan, M. (2017). Causal learning is collaborative: Examining explanation and exploration in social contexts. *Psychonomic Bulletin & Review*, *24*, 1548–1554. https://doi.org/10.3758/s13423-017-1351-3

Loewenstein, J., Thompson, L., & Gentner, D. (2003). Analogical learning in negotiation teams: Comparing cases promotes learning and transfer. *Academy of Management Learning and Education*, *2*, 119–127. https://doi.org/10.5465/amle.2003.9901663

Lombrozo, T. (2006). The structure and function of explanations. *Trends in Cognitive Sciences*, *10*, 464–470. https://doi.org/10.1016/j.tics.2006.08.004

Lombrozo, T. (2016). Explanatory preferences shape learning and inference. *Trends in Cognitive Sciences*, *20*, 748–759. https://doi.org/10.1016/j.tics.2016.08.001

Lombrozo, T., & Carey, S. (2006). Functional explanation and the function of explanation. *Cognition*, *99*, 167–204. https://doi.org/10.1016/j.cognition.2004.12.009

Markman, A., & Gentner, D. (1993). Structural alignment during similarity comparisons. *Cognitive Psychology*, *25*, 431–467. https://doi.org/10.1006/cogp.1993.1011

Mason, L. (2004). Fostering understanding by structural alignment as a route to analogical learning. *Instructional Science*, *32*, 293–318. https://doi.org/10.1023/b:truc.0000026512.88700.32

Matlen, B. J., Gentner, D., & Franconeri, S. L. (2020). Spatial alignment facilitates visual comparison. *Journal of Experimental Psychology*, *46*, 443–457. https://doi.org/10.1037/xhp0000726

Mix, K. S., Smith, L. B., & Crespo, S. (2018). Leveraging relational learning mechanisms to improve place value instruction. In A. Norton, & M. W. Alibali (Eds.), *Research in mathematics education* (pp. 87–121). Springer.

Muldner, K., Burleson, W., & Chi, M. T. (2014). Learning from self-explaining emergent phenomena. In Polman, J. L., Kyza, E.A., O'Neill D. K., Tabak, I., Penuel W. R., Jurow, A. S., O'Connor, K., Lee T., & D'Amico L. (Eds.), *Learning and Becoming in Practice: The International Conference of the Learning Sciences (ICLS) 2014*. (Vol. 2, pp. 847–854). International Society of the Learning Sciences.

Murphy, A. N., Zheng, Y., Shivaram, A., Vollman, E., & Richland, L. E. (2021). Bias and sensitivity to task constraints in spontaneous relational attention. *Journal of Experimental Child Psychology*, *202*, Article e104981. https://doi.org/10.1016/j.jecp.2020.104981

Nokes-Malach, T. J., VanLehn, K., Belenky, D. M., Lichtenstein, M., & Cox, G. (2013). Coordinating principles and examples through analogy and self-explanation. *European Journal of Psychology of Education*, *28*, 1237–1263. https://doi.org/10.1007/s10212-012-0164-z

Paik, J. H., & Mix, K. S. (2006). Preschoolers' use of surface similarity in object comparisons: Taking context into account. *Journal of Experimental Child Psychology*, *95*, 194–214. https://doi.org/10.1016/j.jecp.2006.06.002

R Core Team. (2024). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/. Date accessed March 13, 2026.

Rittle-Johnson, B. (2006). Promoting transfer: Effects of self-explanation and direct instruction. *Child Development*, *77*, 1–15. https://doi.org/10.1111/j.1467-8624.2006.00852.x

Rittle-Johnson, B., Loehr, A. M., & Durkin, K. (2017). Promoting self-explanation to improve mathematics learning: A meta-analysis and instructional design principles. *ZDM*, *49*, 599–611. https://doi.org/10.1007/s11858-017-0834-z

Rittle-Johnson, B., & Star, J. R. (2009). Compared with what? The effects of different comparisons on conceptual knowledge and procedural flexibility for equation solving. *Journal of Educational Psychology*, *101*, 529–544. https://doi.org/10.1037/a0014224

Rottman, B. M., Gentner, D., & Goldwater, M. B. (2012). Causal systems categories: Differences in novice and expert categorization of causal phenomena. *Cognitive Science*, *36*, 919–932. https://doi.org/10.1111/j.1551-6709.2012.01253.x

Sagi, E., Gentner, D., & Lovett, A. (2012). What difference reveals about similarity. *Cognitive Science*, *36*, 1019–1050. https://doi.org/10.1111/j.1551-6709.2012.01250.x

Sidney, P. G., Hattikudur, S., & Alibali, M. W. (2015). How do contrasting cases and self-explanation promote learning? Evidence from fraction division. *Learning and Instruction*, *40*, 29–38. https://doi.org/10.1016/j.learninstruc.2015.07.006

Thompson, C. A., & Opfer, J. E. (2010). How 15 hundred is like 15 cherries: Effect of progressive alignment on representational changes in numerical cognition. *Child Development*, *81*, 1768–1786. https://doi.org/10.1111/j.1467-8624.2010.01509.x

Walker, C. M., Bonawitz, E., & Lombrozo, T. (2017). Effects of explaining on children's preference for simpler hypotheses. *Psychonomic Bulletin & Review*, *24*, 1538–1547. https://doi.org/10.3758/s13423-016-1144-0

Walker, C. M., & Lombrozo, T. (2017). Explaining the moral of the story. *Cognition*, *167*, 266–281. https://doi.org/10.1016/j.cognition.2016.11.007

Walker, C. M., Lombrozo, T., Legare, C. H., & Gopnik, A. (2014). Explaining prompts children to privilege inductively rich properties. *Cognition*, *133*, 343–357. https://doi.org/10.1016/j.cognition.2014.07.008

Wang, S., & Baillargeon, R. (2008). Can infants be "taught" to attend to a new physical variable in an event category? The case of height in covering events. *Cognitive Psychology*, *56*, 284–326. https://doi.org/10.1016/j.cogpsych.2007.06.003

Willard, A. K., Busch, J. T. A., Cullum, K. A., Letourneau, S. M., Sobel, D. M., Callanan, M., & Legare, C. H. (2019). Explain this, explore that: A study of parent–child interaction in a children's museum. *Child Development*, *90*, Article e598–e617. https://doi.org/10.1111/cdev.13232

Williams, J. J., & Lombrozo, T. (2010). The role of explanation in discovery and generalization: Evidence from category learning. *Cognitive Science*, *34*, 776–806. https://doi.org/10.1111/j.1551-6709.2010.01113.x

Zheng, Y., Matlen, B., & Gentner, D. (2022). Spatial alignment facilitates visual comparison in children. *Cognitive Science*, *46*, Article e13182. https://doi.org/10.1111/cogs.13182

Zheng, Y. S., Goldwater, M., & Gentner, D. (2025). Structural alignment and linguistic contrast help children learn a key principle of spatial construction. *Cognitive Science*, *49*, Article e70149. https://doi.org/10.1111/cogs.70149