



Generating explanations via analogical comparison

Christian Hoyos¹ · Dedre Gentner¹

Published online: 24 April 2017
© Psychonomic Society, Inc. 2017

Abstract Generating explanations can be highly effective in promoting learning in both adults and children. Our interest is in the mechanisms that underlie this effect and in whether and how they operate in early learning. In adult reasoning, explanation may call on many subprocesses—including comparison, counterfactual reasoning, and reasoning by exclusion; but it is unlikely that all these processes are available to young children. We propose that one process that may serve both children and adults is comparison. In this study, we asked whether children would use the results of a comparison experience when asked to explain why a model skyscraper was stable. We focused on a challenging principle—that diagonal cross-bracing lends stability to physical structures (Gentner et al., *Cognitive Science*, 40, 224–240, 2016). Six-year-olds either received no training or interacted with model skyscrapers in one of three different conditions, designed to vary in their potential to invite and support comparison. In the Single Model condition, children interacted with a single braced model. In the comparison conditions (Low Alignability and High Alignability), children compared braced and unbraced models. Following experience with the models, children were asked to explain why the braced model was stable. They then received two transfer tasks. We found that children who received highly alignable pairs were most

likely to (a) produce brace-based explanations and (b) transfer the brace principle to a dissimilar context. This provides evidence that children can benefit from analogical comparison in generating explanations and also suggests limitations on this ability.

Keywords Explanation · Comparison · Analogy · Learning

Generating explanations has been shown to be effective in promoting learning (Chi, Bassok, Lewis, Reimann, & Glaser, 1989; Legare & Lombrozo, 2014; Lombrozo, 2010; Siegler, 2002). Yet the processes of explanation generation are largely uncharted—perhaps in part because generating explanations is a complex phenomenon that may draw on many specific processes (Lombrozo, 2012; Williams & Lombrozo, 2010). One key process that has been proposed to be important in generating explanations is analogical comparison. This idea has received theoretical and empirical support in research on adults, as we review below. Our interest here is in examining these processes in children. The study of whether and how young children use comparison to inform their explanations can provide insight into children’s reasoning processes and potentially shed light on the more complex operation of explanatory processes in adults. We begin by reviewing evidence on the power of explanation to promote learning. We then turn to studies that aim to identify the processes that support explanation, focusing particularly on comparison as one such process.

Many studies have shown that learning is improved when learners are prompted to generate an explanation relative to learners who perform a control task (Chi, de Leeuw, Chiu, & LaVancher 1994; Legare & Lombrozo, 2014; Siegler, 2002; Williams & Lombrozo, 2010, 2013; Wong, Lawson, & Keeves, 2002). This is sometimes referred to as the *self-*

This research was supported by NSF Grant SBE-0541957, the Spatial Intelligence and Learning Center (SILC), ONR Grant N00014-16-1-2613 to the second author and an NSF Graduate Research Fellowship DGE-0824162 to the first author.

✉ Christian Hoyos
choyos@u.northwestern.edu

¹ Department of Psychology, Northwestern University, 2029 Sheridan Rd., Evanston, IL 60208, USA

explanation effect, or simply as the *explanation effect* (Chi et al., 1989; Chi et al., 1994). Chi and colleagues propose that generating explanations can lead to new inferences, reveal gaps in understanding, and construct or repair mental models (Chi et al., 1994; Fonseca & Chi, 2011; see also Collins & Gentner, 1987). For example, Williams and Lombrozo (2013) presented adults with multiple items belonging to artificial categories and asked them either to *explain* why each item was a member of the category or to *describe* the items. They found that the explanation group was better at discovering the category rules than the description group. Explanation prompts such as these have been shown to support transfer of knowledge to novel cases in domains as varied as false belief understanding (Amsterlaw & Wellman, 2006), problem solving (Crowley & Siegler, 1999), and algebra (Rittle-Johnson, 2006).

How does explanation support learning? Siegler (2002) proposed that prompts to explain lead learners to search deeply for relevant explanations. Siegler also emphasized that explanation prompts can motivate learners to spend more time with the materials and increase engagement. Lombrozo (2012) proposed that explanation recruits many different cognitive processes, which may include inductive and deductive reasoning, categorization, causal reasoning, and analogy. One specific advantage of explanation, according to Lombrozo's (2012) subsumptive constraints account, is that when generating explanations, people seek general principles that unify multiple examples. The identification of a single unifying pattern supports generalization and helps learners move away from surface features. In addition, engaging in explanation prompts a search for information that can prepare learners for future contexts (Lombrozo & Carey, 2006).

Many different subprocesses—including comparison, inductive and deductive reasoning, counterfactual reasoning, causal reasoning, and memory retrieval—have been proposed to enter into the generation of explanations. For instance, inherent features—which are more easily accessible in memory—tend to be used in explanations (Cimpian, 2015; Cimpian & Salomon, 2014). Novel evidence that conflicts with prior beliefs can influence how learners explore and identify phenomena that call for explanations (Bonawitz, van Schijndel, Friel, & Schulz, 2012; Legare, 2012, 2014; Legare, Gelman, & Wellman, 2010; Williams & Lombrozo, 2013). Our focus here is on comparison processes. When learners engage in explanation, they often carry out comparisons among exemplars that help identify what needs to be explained and give rise to information that can be used in their explanations. Our question is whether children are able to use these kinds of comparison processes in service of generating explanations.

The idea that comparison is involved in generating explanation has been explored by philosophers and cognitive psychologists (Chin-Parker & Bradner, 2010; Hilton & Slugoski, 1986; Hitchcock & Knobe, 2009; Kahneman & Miller, 1986;

Landy & Hummel, 2010; Phillips, Luguri, & Knobe, 2015; van Fraassen, 1980; Weiner, 1985). The common theme across many of these approaches is that generating explanations involves calling forth a *contrast class* with which the target phenomenon is compared. Comparing the target with the contrast class can serve at least two distinct functions: identifying the *explanandum* (i.e., the phenomenon that needs to be explained) and finding or generating the *explanans* (i.e., the information that does the explaining). For instance, Chin-Parker and Bradner (2010) proposed that the first step in answering a “why” question involves implicitly comparing the question to other potential questions in order to define what exactly needs to be answered. For example, in order to answer the question “Why did you order the cheese melt?”, one must compare it to other possible alternatives (such as other menu options) that will guide the explainer toward the relevant topic. In the same vein, others have suggested that abnormal events are implicitly contrasted with normal events in order to identify both *explananda* and *explanantia* (Hilton & Slugoski, 1986; Hitchcock & Knobe, 2009; Kahneman & Miller, 1986; Weiner, 1985).

Further evidence for the role of comparison in explanation comes from studies that ask learners to generate explanations in the presence of multiple exemplars (rather than to retrieve cases from memory, as in the studies just discussed; Edwards, Williams, & Lombrozo, 2013; Edwards, Williams, Lombrozo, & Gentner, under review; Nokes-Malach, VanLehn, Belenky, Lichtenstein, & Cox, 2013; Renkl, 2014; Richey, Zepeda, & Nokes-Malach, 2015; Sidney, Hattikudur, & Alibali, 2015). One outcome of this research is the observation that self-explanation and analogical comparison have important commonalities. Both have been found to support the acquisition of abstract knowledge and the discovery of deep commonalities across instances (Richey & Nokes-Malach, 2015). Indeed, both activities have been described as *constructive* (Chi, 2009), or by the phrase, “learning by thinking,” (T. Lombrozo, personal communication, as cited in Xu, 2016), because they help learners generate new knowledge.

Beyond these parallels, some recent studies have investigated the possibility of a more active interaction between the two processes: specifically, the possibility that explanation can call on comparison. For example, Edwards et al. (2017), gave adults a category learning task and prompted them either to compare exemplars of the categories or to explain why a given exemplar belonged in its category. After completing the classification task, people reported (on a 1–7 scale) the extent to which they had noticed themselves comparing exemplars. Strikingly, the results showed that prompting participants to explain led them to report more comparison than did prompting them to compare (see also Edwards et al., 2013). In a similar vein, Sidney et al. (2015) examined explanation and analogical comparison both separately and together in undergraduates' understanding of fraction division and found

that an explanation prompt supported greater conceptual understanding than contrasting cases alone. However, self-explanation also led learners to report noticing more similarities and differences across cases than did those who did not self-explain—consistent with the results found by Edwards et al.

How might comparison subserve explanation? The prior studies have led to at least three proposals for how comparison is involved in explanation: (1) a comparison between expected and unexpected outcomes can spontaneously elicit a search for an explanation (e.g., Legare et al., 2010; Weiner, 1985), (2) comparisons between a target phenomenon and other possible alternatives can help learners identify what needs to be explained, and (3) comparison across exemplars can reveal commonalities or differences that learners include in their explanations (Chin-Parker & Bradner, 2010; Edwards et al., 2017; Hilton & Slugoski, 1986; Hitchcock & Knobe, 2009; Kahneman & Miller, 1986; Landy & Hummel, 2010; Sidney et al., 2015). In the current work, we focus chiefly on the third proposal and ask whether and how young children use comparison to inform their explanations. Many of these studies have focused on adult participants who have a vast store of knowledge from which cases can be drawn. By focusing on children, who have fewer resources than adults, we can better understand how the comparison process is involved in generating explanations. Before turning to our experiment, we first describe how analogical comparison works and how it supports learning.

Analogical comparison—comparison that involves aligning relational structure as well as noting matching features—can act as a learning process that can support both inference and abstraction. Comparing across exemplars has been shown to benefit learning and transfer in business school students learning negotiation strategies (Gentner, Loewenstein, & Thompson, 2003), college students learning categories (Higgins & Ross, 2011) or principles (Gick & Holyoak, 1983; Kurtz, Miao, & Gentner, 2001), middle-school and high-school students learning mathematics and science (Richland, Zur, & Holyoak, 2007; Rittle-Johnson & Star, 2007, 2009; Schwartz, Chase, Oppezzo, & Chin, 2011), and preschoolers learning novel words and relations (Augier & Thibaut, 2013; Christie & Gentner, 2010; Gentner & Namy, 1999, 2006; Kotovsky & Gentner, 1996; see also Goldstone, Day, & Son, 2010, for a review, and Alfieri, Nokes-Malach, & Schunn, 2013, for a meta-analysis).

Using Gentner's (1983) structure-mapping theory as a framework, we describe analogical comparison as a process of structural alignment and inference (Falkenhainer, Forbus, & Gentner, 1989; Gentner, 1983; Gentner & Markman, 1997). The process of structural alignment leads to learning in several ways: it can reveal commonalities and differences (Gentner & Markman, 1994; Markman & Gentner, 1996; Sagi, Gentner, & Lovett, 2012), lead to new inferences (Clement & Gentner,

1991; Day & Gentner, 2007; Markman, 1997; Spellman & Holyoak, 1996), and give rise to abstract commonalities while deemphasizing nonmatching surface features (Christie & Gentner, 2010; Doumas & Hummel, 2013; Gentner & Namy, 1999, 2006; Gick & Holyoak, 1983; Kotovsky & Gentner, 1996). Further, structural alignment often serves to make previously implicit relations more salient for the learner. A process model of structure-mapping has been implemented in SME, the Structure-Mapping Engine (Falkenhainer et al., 1989; Forbus, Gentner & Law, 1995; Forbus, Ferguson, Lovett, & Gentner, 2016).

A further effect of structural alignment—and one highly relevant to the present work—is that alignable differences often emerge as a natural outcome (Gentner & Markman, 1994; Markman & Gentner, 1993, 1996; Sagi et al., 2012). *Alignable differences* are differences that share the same role within each of the aligned relational structures. For example, Markman and Gentner (1996) found that when two similar figures were presented to participants, they were able to list more differences (primarily alignable differences) than when low-similarity figures were presented; the same pattern held for pairs of concepts presented verbally (Gentner & Markman, 1994; Markman & Gentner, 1993). Likewise, Sagi et al. (2012) found that participants were faster in stating differences between high-similarity pairs of figures than between low-similarity pairs. Thus, structural alignment highlights differences connected to the common relational structure. The similarity of the pair matters here for two reasons: first, pairs that are high in overall similarity (both relational similarity and surface similarity) are highly likely to engage spontaneous comparison processes, and, second, such pairs are easier and faster to align than low-similarity pairs. This is the case even if the low-similarity pair shares the same relational structure, because in high-similarity pairs the obvious object matches support the relational alignment (Gentner & Kurtz, 2006; Gentner & Toupin, 1986). This alignability advantage for difference detection is a key signature of structural alignment.

The above findings suggest a way to help learners notice subtle but important features of a situation—namely, by comparing alignable pairs designed so that the key feature appears as an alignable difference. In the current study, we aimed to teach children a basic engineering principle—the idea that within a quadrilateral, a diagonal brace confers structural stability.¹ We know from prior research that diagonal elements are less salient to young children than are horizontal and vertical elements (Olson, 1970). Our question is whether children will engage in comparison when asked to explain what makes a building strong. If so, they should produce better explanations, because the alignable difference (the key feature of a

¹ A diagonal brace within a quadrilateral prevents it from being deformed into a rhombus. This follows from the fact that a triangle (unlike a quadrilateral) is a stable polygon.

brace) will be highlighted and therefore more likely to be included in their explanations. A further question is whether this knowledge will transfer to other situations. We propose that facilitating comparison will also help children transfer knowledge.

We begin by reviewing a closely related prior study that sets the stage for the present study. When constructing toy buildings out of an erector set, children often neglect the important role of diagonal cross-bracing, and generally only use vertical and horizontal pieces in their models (Benjamin, Haden, & Wilkerson, 2010). Prior work has shown that facilitating comparison between contrasting cases can support learning the brace principle (Gentner et al., 2016). In this study, 6-year-old children were asked to compare model skyscrapers. In one condition (High Alignability) the training models were highly alignable (they largely shared common internal structure) and shared high overall similarity (they had many matching parts). In another condition (Low Alignability), the models also shared common structure, but were less similar overall (they contained some dissimilar parts; see Fig. 1). In both pairs, the two buildings differed in that one of them had a brace (and therefore had stable structure) and the other had only horizontal and vertical pieces (and therefore could be bent sidewise). A third group of children received no training. During training, children were asked ‘Which is stronger?’² After wiggling each of the models, children recognized that the braced building was stronger. Importantly, the brace was never pointed out to the child. Later, children were individually given a repair task, in which they had add a new piece to another building to make it strong (described in more detail below). Children in the High Alignability condition produced more diagonal braces than did children in the Low Alignability and No Training conditions—suggesting that they had carried out a structural alignment, resulting in pop-out of the brace as an alignable difference (Gentner et al., 2016).

In the current research, we adapt this method to ask whether 6-year-old children will use comparison when asked to explain. We compared four conditions. Two of them were the same as in the study just described: High Alignability (high overall similarity, relationally alignable pairs) and Low Alignability (low overall similarity, relationally alignable pairs); in both cases, the brace should emerge as an alignable difference if children carry out a full structural alignment. The other two conditions were Single Model (a braced model), and No Training. In contrast to the prior study, here we ask children to explain why the braced building is strong(er). If children compare the two buildings and use the results of that comparison in their explanations, then the High Alignability

group should produce more brace-based explanations than the Single Model group. This is because the highly alignable pair both invites comparison and makes it easy to arrive at a structural alignment of the models, resulting in the key alignable difference ‘popping out.’ If children in the Low Alignability condition also use the results of their comparison to inform their explanations, then we may also see more brace-based explanations in this group than in the Single Model group. However, because the process of structural alignment should be less fluent in the Low Alignability condition, we expect fewer brace-based explanations in the Low Alignability condition than in the High Alignability condition.

In addition to examining the content of children’s explanations, we also tested children in transfer tasks to see how well they could apply the brace principle. We used both a near transfer task (using another model building, as in the original museum study) and a far transfer task (a motorcycle) to test whether children could transfer the brace principle to novel contexts.

To review, we predict that children in the High Alignability group will produce more brace-based explanations and perform better on the transfer task than will the Single Model group. We further expect the High Alignability group to show an advantage over the Low Alignability group. Whether the Low Alignability group will differ from the Single Model group is an open question. On the one hand, they do have the opportunity to compare, but on the other hand, achieving a full alignment should be rather challenging.

Method

Participants

A total of 72 six-year-olds participated in the study ($M = 77.7$ months, range: 6.0–7.0; 35 male), divided into four training conditions, each with 18 children: No Training (NT), Single Model (SM), Low Alignability (LA), and High Alignability (HA). Families were recruited from the Evanston/Chicago area and the racial/economic composition of the sample reflected that of the local population (majority European American, middle and upper middle class). Children received a small gift for their participation.

Materials

During the training phase, children either saw a single braced model skyscraper (SM) or two model skyscrapers (LA or HA; see Fig. 1). These models were constructed from a custom-made erector set, similar to the models used in Gentner et al. (2016). The models were 2 feet tall and 13 inches on each side. The near transfer task was also adapted from Gentner et al. In this task, children had to repair a 1-foot-high model cube using

² We used the term *strong* instead of the more correct term *stable* to facilitate children’s understanding. The unbraced building could be bent to the side, whereas the braced model remained upright.

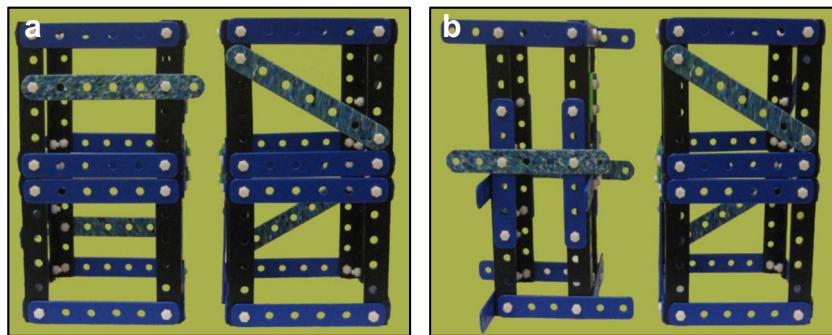


Fig. 1 Model skyscrapers shown during training. **a** High Alignability pair. **b** Low Alignability pair. For the Single Model condition, only the braced model on the right side was shown to children

an extra beam (see Fig. 2). For the far transfer task, a small motorcycle was constructed with a Stanley Construct & Play set. A square shape was embedded inside the motorcycle, and this was the target area that children were asked to repair with an additional piece. Finally, between the training and transfer tasks, were two filler tasks. In the first filler task, children saw images of four patterns for 12 trials and had to select which among the patterns was different from the others. The second filler task was made up of nine trials of a modified version of Ravens' matrices. Each filler task took approximately 5 minutes to complete.

Procedure

Training For the training phase, children were placed in one of four training conditions: NT, SM, LA, and HA. In the NT condition, children did not interact with model skyscrapers and completed only the filler tasks and repair tasks. This group served as a baseline.

In all conditions, the presence of diagonal bracing was never pointed out to children. In the SM condition, children were shown a single braced model. Children were asked whether they thought the “building was strong,” and then invited to test the model’s stability by wiggling it, (which revealed that the model remained upright). Children were then asked to generate an explanation for the stability of the model;

they were asked, “Why do you think this building is strong?” If children were reluctant to produce an explanation, or if they said “I don’t know,” the same question was asked again and they were given time to look at the model. If after an additional prompt children still failed to produce an explanation, the experimenter then moved on to the first filler task. Before the filler task, the experimenter removed the models from the table, and they were placed out of the child’s sight for the remainder of the study.

In the LA and HA conditions, children were shown two model skyscrapers (see Fig. 1) and asked to guess (before they touched the models), which of the two was “stronger.” Children guessed randomly between the two, confirming that the diagonal is not readily obvious to 6-year-olds. They were then invited to wiggle each model. They saw that the braced model did not move, but the unbraced model shook substantially and could be bent over. All children then identified the braced model as the stronger one and were asked to generate an explanation for its stability; they were asked “Why do you think this building is stronger than this one?” After the child produced an explanation, the models were removed from sight and children completed the ensuing tasks.

Transfer tasks After the training phase, children completed the first filler task. Following this, children completed the near transfer task (see Fig. 2). In this task, a model cube was placed



Fig. 2 A child completing the near transfer task. The model is first demonstrated to be “wobbly.” Then the child is given a new piece and asked to repair the model. This shows a child placing the new piece in a diagonal orientation

on the table, and children were told that the experimenter's friend created the building but that it was "wobbly." The experimenter shook the cube to demonstrate that it could be completely distorted from its original shape. The child was then presented with an additional beam and was asked, "Where should we put this piece to make the building strong?" The experimenter recorded the orientation of the piece—whether the child placed the beam horizontally, vertically, or diagonally. The piece was considered a diagonal if it was tilted off the horizontal axis by at least one notch on the model. Once the child made a selection, the experimenter thanked the child and moved onto the next task (the beam was not actually screwed onto the model cube so that children did not receive feedback regarding their selection).

After the near transfer task, children completed the second filler task, followed by the far transfer task. The small motorcycle was shown to children, and the experimenter demonstrated that the square part could be distorted and bent from side to side. The child was then presented with an additional beam and was asked to repair the motorcycle. As before, the orientation of the piece was recorded. This marked the end of the session.

Results

Explanations

Children produced explanations about structural stability after interacting with the model skyscraper(s). The explanations were transcribed and presented (with no information as to condition) to the first author and a rater who was blind to the hypotheses of the study. The raters categorized the explanations into nine separate categories, provided in Table 1 along with examples of each type. Cohen's kappa was used to calculate interrater reliability, and this showed a substantial level of agreement ($\kappa = .78$). Disagreement was resolved by discussion. None of the brace-based explanations were disagreed upon. These results are shown in Table 1.

As predicted, children in the HA condition ($n = 9$) were most likely to produce brace-based explanations. In contrast, no children in the SM ($n = 0$) condition used the brace in their explanations, and only a few children in the LA condition did so ($n = 3$). Chi-square tests were used to calculate differences among the groups and showed that children in the HA condition referred to the brace in their explanations significantly more often than children in the LA and SM conditions, $\chi^2(1, N = 36) = 4.5, p = .03, \chi^2(1, N = 36) = 12, p = .001$, respectively.³ Our second question was whether the LA group would differ from the SM group. It appears that they did. As shown

in Table 1, the SM group mostly referred to intrinsic features of the model (strong material, tight screws, etc.). Although some children within the LA group also gave such responses, four of them referred to shape differences in their explanations. Combining these responses with their three brace-based explanations, we see that the LA group ($n = 7$) gave significantly more comparison-based explanations than the SM group ($n = 0$), $\chi^2(1, N = 36) = 8.7, p = .003$.

Figure 3 shows a subset of these data, grouped for ease of comparison across conditions. We excluded categories used by fewer than three participants (quantity, creation, none); we also excluded the tautology category, which had roughly equal numbers across groups. We also grouped together three categories (construction, screws, material) that referred to nondiagnostic features that were shared across all models. Figure 3 shows that, as expected, children in the single model condition, lacking any comparison opportunity, focused largely on nondiagnostic features of the single building. Although some children in the LA condition also gave nondiagnostic features, many of them reported differences between the models. And finally, children in the HA condition were the most likely to include the brace in their explanations.

Transfer tasks

Performance on the transfer tasks is shown in Fig. 4. For the near transfer task, the orientation of the beam was recorded (diagonally or nondiagonally). Logistic regression was used to calculate the likelihood that children would produce a diagonal. We used condition as a predictor, and the NT group served as the reference group. We found that the proportion of children producing diagonals in the HA ($M = 0.61$), LA ($M = 0.56$), and SM conditions ($M = 0.78$), was significantly greater than the proportion producing diagonals in the NT group ($M = .11$). The results of this analysis are shown in Table 2.

Likewise, for the far transfer task, children's diagonal placements were recorded and logistic regression was used to calculate the likelihood that children produced a diagonal to repair the model. Condition was used as the predictor. We found that the proportion of children producing diagonals in the HA condition ($M = 0.67$) was significantly greater than for the NT group ($M = 0.11$). Neither the LA ($M = 0.39$) nor the SM ($M = 0.39$) conditions were significantly different from the NT condition (see Table 2).

We explored whether there were different rates of diagonal production in the transfer tasks based on children's production of brace-based explanations. Across all conditions, only 12 children included the brace in their explanations. Of these children, eight produced a brace in the near transfer task, and six produced a brace in the far transfer task. However, we did not find production of brace-based explanations to be a significant predictor of brace production in either the near transfer task, $\beta = .11, SE = 0.69$, Wald's $\chi^2 = .02, p =$

³ A logistic regression could not be used for this analysis. Because no children in the single model group produced brace-based explanations, the model would not converge.

Table 1 Categories of explanations produced by children

Category	Number by Condition			Examples
	SM	LA	HA	
Brace	0	3	9	“The diagonal is stronger than this piece.” “Over here it’s straight, and over here, it’s like that.” [diagonal gesture]
Shape	0	4	0	“It has more features and is wider.” “It’s thicker.”
Construction	2	2	0	“It has more put together parts.” “It has a big square at the bottom and its screwed on.”
Screws	7	2	4	“The bolts are a little tighter.” “It has a lot of screws.”
Material	5	1	0	“It’s made of plastic, strong plastic.” “It’s made out of strong stuff.”
Quantity	0	2	0	“It has more blocks.” “It has more blocks than that one.”
Creation	0	0	1	“They hammered it better.”
Tautology	4	4	3	“It’s stable and stronger than the other—the other is wobbly.” “This one is stronger.”
None	0	0	1	“I don’t know.”

Note. SM = Single Model; LA = Low Alignability; HA = High Alignability

.88, or the far transfer task, $\beta = .01$, $SE = 0.66$, Wald’s $\chi^2 = .02$, $p = .88$.

Discussion

The goal of this study was to examine the role of analogical comparison in explanation—specifically, whether children can use comparison to inform their explanations. We gave 6-year-old children a challenging task—to explain why a model building was “strong”—that is, why it maintained a stable

structure when pushed. The key point is that the building has a diagonal piece which acts as a brace. Consistent with prior evidence that diagonal elements are not understood by young children (Benjamin et al., 2010; Gentner et al., 2016; Olson, 1970), we found that none of the children who saw a single braced model mentioned the diagonal brace; instead, they focused on intrinsic features, such as the material the building was made of. This set the stage for asking whether children will exploit a comparison if one is available.

If children can make use of comparisons in producing explanations, then children who have an opportunity to compare should produce different explanations than those who do not. More specifically, the outcome of the comparison process should influence what information enters into the explanation. We tested these claims by varying whether children had available comparisons, and (to test the second claim) by varying

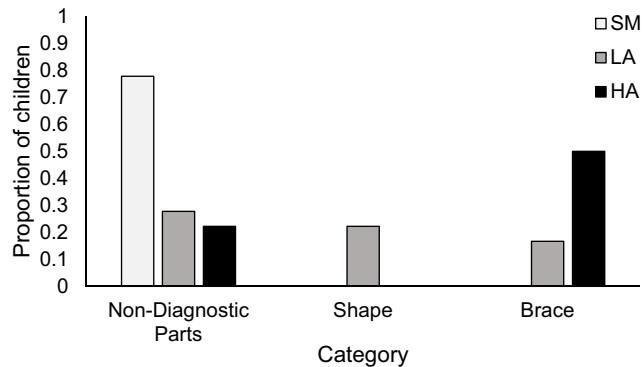


Fig. 3 Major patterns of explanation produced by children across conditions. Each bar is the proportion of children producing that explanation type within each condition. Nondiagnostic parts consists of three categories: construction, screws, and material. Children in the HA condition were more likely to include the brace in their explanations than were those in the other conditions. SM = Single Model; LA = Low Alignability; HA = High Alignability

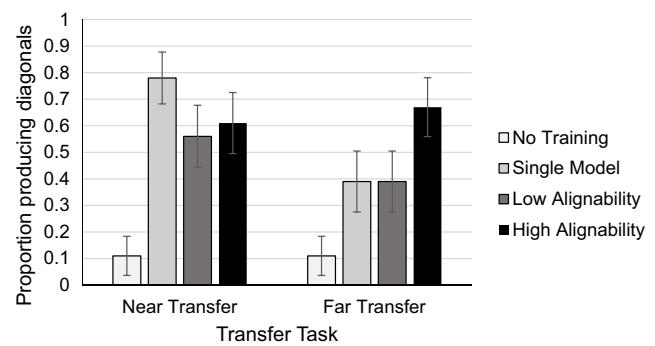


Fig. 4 Proportion of children producing diagonals in both near and far transfer tasks. Error bars depict ± 1 SE

Table 2 Logistic regression predicting children's diagonal production by condition

Task	Predictor	β	SE	Wald's χ^2	<i>p</i>	Odds ratio
Near transfer	HA	2.53	0.89	8.05	.005	12.57
	LA	2.30	0.89	6.73	.009	10.00
	SM	3.33	0.94	12.56	<.001	28.00
Far transfer	HA	2.77	0.90	9.46	.002	16.00
	LA	1.63	0.89	3.33	.068	5.09
	SM	1.63	0.89	3.33	.068	5.09

Note. HA = High Alignability; LA = Low Alignability; SM = Single Model. Performance in each condition is compared against the baseline No Training group. Boldface *p* values denote significance. Comparisons from the LA and SM conditions in the far transfer task are identical because they have the same mean and sample size. Children across all three conditions produced more braces than the No Training group in the near transfer task, but children in the HA condition produced significantly more braces than the no No Training in the far transfer task

the alignability of the comparison available to children. We found support for both claims. First, children in the High and Low Alignability groups produced different types of explanations than children in the Single Model group. This suggests that children made use of comparison by focusing on differences between the models for their explanations. Further, we found that children in the High Alignability group, who received ideal conditions for structural alignment, produced significantly more brace-based explanations than both the Single Model and Low Alignability groups. The highly alignable pair both invited and supported comparison, and this led children to notice an alignable difference—the presence of diagonal bracing—that they then used in their explanations.

A further prediction was that facilitating comparison would support transfer. Consistent with this prediction, the High Alignability group performed significantly better than the No Training group in the far transfer task. (We discuss the near transfer task later).

In addition, we posed an interesting open question concerning the Low Alignability group. We know from prior research that spontaneous comparison is unlikely for children given this low-similarity pair. Thus, if children in this group carry out a comparison and use the results in their explanations, this will suggest that 6-year-old children share—at least to some extent—the adult intuition that comparison is useful in explanation. Indeed, we found that the Low Alignability group gave more comparison-based explanations than the Single Model group. These were roughly divided between brace-based explanations (indicating a full structural alignment of the internal structures of the models) and shape-based explanations (suggesting that these children simply compared the external shapes of the two models). Thus, although the Low Alignability group was not as successful as the High Alignability group in deriving the brace principle, the results suggest that they did use comparison to inform their explanations.

These results are consistent with Gentner et al.'s (2016) findings in the Chicago Children's Museum. As in the current study, children given a highly alignable pair were likely to notice the brace and to use it in a near transfer task. The current study extends this work in showing that 6-year-old children will use the results of a comparison task to support the generation of explanations and can go on to complete a more difficult transfer task. Even children given low-alignability pairs were likely to carry out a comparison and use the results in their explanation. They were less likely to arrive at the ideal explanation, but the fact remains that they used comparison to derive explanatory insight.

One puzzling result is that all three groups performed significantly better than the No Training group in the near transfer task. We had expected the Single Model group to perform similarly to the No Training group, since children had no basis on which to pick out the brace from the model. In retrospect, we suspect that their success was due to the close resemblance of the near transfer task materials to the training materials. Children in the Single Model group may have recalled the model they saw during training and merely copied the prior model, without noting its relevance for stability. Consistent with this account, their performance fell on the far transfer task.

A second puzzling finding is that we did not find links between verbally producing brace-based explanations and performance on the far transfer task. Overall, very few children (a total of 12 out of 54) produced a brace-based explanation. Only half of these children produced braces in the far transfer task. Further, some children who did not produce a brace-based explanation still went on to produce braces in the far transfer task. The first discrepancy—that half the children who referred to the diagonal brace in their explanations failed to use a brace in the far transfer task—fits with other findings suggesting that transferring a known principle to a very different context is difficult. Consistent with this possibility, children did somewhat better at transferring to the near transfer task, which resembled the training task; eight of the 12 children who mentioned a brace in their explanation went on to use it in the near transfer task, compared to the six out of 12 that produced a brace in the far transfer task. We suspect that the second discrepancy, that some children who failed to mention the brace in their explanations nonetheless produced them in the transfer tasks, may simply be an instance of children's verbalizable knowledge lagging behind their spatial-perceptual knowledge. On this account, the children who produced a brace-based explanation may underrepresent those who gained insight from the training.

Relation to prior work

Prior work has shown that the search for explanations is influenced by the interaction between prior knowledge and novel

evidence (Bonawitz et al., 2012; Legare, 2012, 2014; Legare et al., 2010; Williams & Lombrozo, 2013). In the present study, presenting children with contrasting cases highlighted an alignable difference that allowed them to move away from their prepotent responses based on materials. These findings are consistent with the idea that children can benefit from analogical comparison to generate new hypotheses about spatial relational structure (Christie & Gentner, 2010; Gentner & Hoyos 2017; Gentner & Medina, 1998; Xu, 2016).

Previous work examining learning from exemplars has provided some indirect evidence that learners use available comparisons in their explanations. For example, Rittle-Johnson and Star (2007) had middle-schoolers explain similarities and differences across solutions to algebra problems, or consider each problem sequentially. The authors found that the comparison group showed greater flexibility and transfer of knowledge than the sequential group. They also found that the comparison group included more comparisons of the solution methods in their explanations than the sequential group. However, the explanation prompts differed across these groups, so it is unclear whether this was an effect of the prompt or the availability of comparison. The current work provides more direct evidence that available comparisons are used by children in their explanations. Further we reveal one important factor—degree of alignability—that influences how children engage in the comparison process, and the type of information that may emerge from the comparison. Specifically, a high degree of alignability (1) increases the likelihood that structural alignment is initiated and (2) renders the structural alignment process more fluent.

We also reviewed evidence that adults asked to generate explanations reported greater engagement in comparison than those who received explicit instructions to compare (Edwards et al., 2017; Sidney et al., 2015). However, to our knowledge the current study is the first research to explore whether young children recruit comparison when asked to explain. The current work suggests that children also exploit available comparisons to generate explanations. However, an open question is whether they would actively seek such comparisons without explicit cues to compare (an issue which we consider in further detail below).

Revisiting the three proposals

We reviewed three proposals in the introduction regarding the links between comparison and explanation. These included the idea that (1) comparison can spontaneously elicit a search for explanations (Legare et al., 2010; Weiner, 1985); (2) comparing exemplars can help narrow down the explanandum, or what needs to be explained; and (3) comparing exemplars can reveal information to be used in the explanans, or the content of the explanation (Chin-Parker & Bradner, 2010; Edwards et al., 2017; Hilton & Slugoski, 1986; Hitchcock & Knobe,

2009; Kahneman & Miller, 1986; Landy & Hummel, 2010; Sidney et al., 2015). This study focused chiefly on the third proposal—that explanation draws on the products of comparison. It is possible that comparison also influenced the explanandum—that is, that the comparison process may clarify what needs to be explained.

What we can say is that, when comparison pairs were readily available, 6-year-old children used the results of their comparison in their explanation. When this was an ideal, high alignable comparison pair that was readily alignable to reveal a key alignable difference, they were likely to use that difference in their explanation. But even when they received a less alignable pair, they still used the results in their explanation. The Low Alignability group was significantly less likely to produce material-based explanations than were children in the Single Model condition. However, although some children in the Low Alignability group were able to align the internal structure of the models and thereby the brace, others simply noted the overall shape differences between the models.

Future directions

An interesting question is whether children actively seek out comparisons when asked to explain, as adults have been found to do (Edwards et al., 2017; Sidney et al., 2015). This study cannot answer this question, because children in the two comparison conditions were asked “Which one is stronger” in the presence of pairs—thus encouraging them to compare the pairs. In future research, it would be interesting to investigate the degree to which children actively seek out comparison cases when asked to explain. For example, children could be asked why a given building is strong (or wobbly) in a context containing several potential comparison objects. Would children seek out comparison cases, and if so, what would be their basis for selection—surface similarity, or deeper structural features? Further, how would this behavior change developmentally? It seems likely that older and/or more knowledgeable children will seek out comparisons to a greater extent than younger children, and will be more discerning in which comparisons to use.

A further developmental question is whether and when children will generate or retrieve examples from memory when asked to explain in the absence of any physically available comparison objects. A further, related question is whether children are more likely to retrieve comparison cases from memory when asked to explain an abnormal situation. As Kahneman and Miller (1986) noted, “an abnormal event is one that has highly available alternatives, whether retrieved or constructed; a normal event mainly evokes representations that resemble it” (p. 137). Following this reasoning, if children were given a nonbraced building and asked “Why is this one wobbly?”, might they retrieve stable buildings from

memory? Again, we suspect that the ability to do this would increase over development and over gains in knowledge.

These findings have implications for education. The presence of highly alignable exemplars allows children to grasp differences between them that can then be generalized and can be applied in novel contexts. Much research supports the use of contrastive cases in educational contexts (Richland et al., 2007; Rittle-Johnson & Star, 2011; see Alfieri et al., 2013, for a meta-analysis). If, as we suspect, asking for explanations invites children to seek and use the outcomes of comparisons, then explanation can work in tandem with analogical comparison to aid children's learning.

Conclusion

Both explanation and analogical comparison have been found to support learning, and some accounts have posited that comparison is central to explanation. Here we have provided further support for this claim by showing that children can use the results of an available comparison in their explanations. Future work should examine the developmental trajectory of when and how children seek and use comparisons in the service of explanation.

References

Alfieri, L., Nokes-Malach, T. J., & Schunn, C. D. (2013). Learning through case comparisons: A meta-analytic review. *Educational Psychologist*, 48(2), 87–113.

Amsterlaw, J., & Wellman, H. M. (2006). Theories of mind in transition: A microgenetic study of the development of false belief understanding. *Journal of Cognition and Development*, 7(2), 139–172.

Augier, L., & Thibaut, J. P. (2013). The benefits and costs of comparisons in a novel object categorization task: Interactions with development. *Psychonomic Bulletin & Review*, 20(2), 1126–1132.

Benjamin, N., Haden, C. A., & Wilkerson, E. (2010). Enhancing building, conversation, and learning through caregiver-child interactions in a children's museum. *Developmental Psychology*, 46(2), 502–515.

Bonawitz, E. B., van Schijndel, T., Friel, D., & Schulz, L. (2012). Balancing theories and evidence in children's exploration, explanations, and learning. *Cognitive Psychology*, 64(4), 215–234.

Chi, M. T. H. (2009). Active-constructive-interactive: A conceptual framework for differentiating learning activities. *Topics in Cognitive Science*, 1, 73–105.

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

Chin-Parker, S., & Bradner, A. (2010). Background shifts affect explanatory style: How a pragmatic theory of explanation accounts for background effects in the generation of explanations. *Cognitive Processing*, 11, 227–249.

Christie, S., & Gentner, D. (2010). Where hypotheses come from: Learning new relations by structural alignment. *Journal of Cognition and Development*, 11(3), 356–373.

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

Cimpian, A., & Salomon, E. (2014). The inference heuristic: An intuitive means of making sense of the world, and a potential precursor to psychological essentialism. *Behavioral and Brain Sciences*, 37(5), 461–480.

Clement, C. A., & Gentner, D. (1991). Systematicity as a selection constraint in analogical mapping. *Cognitive Science*, 15(1), 89–132.

Collins, A., & Gentner, D. (1987). How people construct mental models. In D. Holland & N. Quinn (Eds.), *Cultural models in language and thought* (pp. 243–265). Cambridge: Cambridge University Press.

Crowley, K., & Siegler, R. S. (1999). Explanation and generalization in young children's strategy learning. *Child Development*, 70, 304–316.

Day, S., & Gentner, D. (2007). Nonintentional analogical inference in text comprehension. *Memory and Cognition*, 35, 39–49.

Doumas, L. A. A., & Hummel, J. E. (2013). Comparison and mapping facilitate relation discovery and predication. *PLOS ONE*, 8(6), e63889. doi:10.1371/journal.pone.0063889

Edwards, B. J., Williams, J. J., & Lombrozo, T. (2013). Effects of explanation and comparison on category learning. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th Annual Conference of the Cognitive Science Society*. Austin: Cognitive Science Society.

Falkenhainer, B., Forbus, K. D., & Gentner, D. (1989). The structure-mapping engine: Algorithm and examples. *Artificial Intelligence*, 41, 1–63.

Fonseca, B. A., & Chi, M. T. H. (2011). Instruction based on self-explanation. In R. E. Mayer & P. A. Alexander (Eds.), *The handbook of research on learning and instruction* (pp. 296–321). New York: Taylor & Francis.

Forbus, K. D., Ferguson, R. W., Lovett, A., & Gentner, D. (2016). Extending SME to handle large-scale cognitive modeling. *Cognitive Science*. doi:10.1111/cogs.12377

Forbus, K. D., Gentner, D., & Law, K. (1995). MAC/FAC: A model of similarity-based retrieval. *Cognitive Science*, 19, 141–205.

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science*, 7, 155–170.

Gentner, D., & Hoyos, C. (2017). Analogy and abstraction. *Topics in Cognitive Science*.

Gentner, D., & Kurtz, K. (2006). Relations, objects, and the composition of analogies. *Cognitive Science*, 30, 609–642.

Gentner, D., Levine, S. C., Dhillon, S., Ping, R., Bradley, C., Isaia, A., & Honke, G. (2016). Rapid learning in a children's museum via analogical comparison. *Cognitive Science*, 40, 224–240.

Gentner, D., Loewenstein, J., & Thompson, L. (2003). Learning and transfer: A general role for analogical encoding. *Journal of Educational Psychology*, 95(2), 393–408.

Gentner, D., & Markman, A. B. (1994). Structural alignment in comparison: No difference without similarity. *Psychological Science*, 5(3), 152–158.

Gentner, D., & Markman, A. (1997). Structure mapping in analogy and similarity. *American Psychologist*, 52, 45–56.

Gentner, D., & Medina, J. (1998). Similarity and the development of rules. *Cognition*, 65, 263–297.

Gentner, D., & Namy, L. (1999). Comparison in the development of categories. *Cognitive Development*, 14, 487–513.

Gentner, D., & Namy, L. (2006). Analogical processes in language learning. *Current Directions in Psychological Science*, 15(6), 297–301.

Gentner, D., & Toupin, C. (1986). Systematicity and surface similarity in the development of analogy. *Cognitive Science*, 10, 277–300.

Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology*, 15(1), 1–38.

Goldstone, R. L., Day, S., & Son, J. Y. (2010). Comparison. In B. Glatzeder, V. Goel, & A. von Müller (Eds.), *Towards a theory of thinking: Building blocks for a conceptual framework* (pp. 103–122). Berlin: Springer.

Higgins, E. J., & Ross, B. H. (2011). Comparisons in category learning: How best to compare for what. In L. Carlson, C. Holscher, & T. Shipley (Eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society*. Austin: Cognitive Science Society.

Hilton, D. J., & Slugoski, B. R. (1986). Knowledge-based causal attribution: The abnormal conditions focus model. *Psychological Review*, 93(1), 75–88.

Hitchcock, C., & Knobe, J. (2009). Cause and norm. *Journal of Philosophy*, 11(11), 587–612.

Kahneman, D., & Miller, D. T. (1986). Norm theory: Comparing reality to its alternatives. *Psychological Review*, 93(2), 136–153.

Kotovsky, K., & Gentner, D. (1996). Comparison and categorization in the development of relational similarity. *Child Development*, 67, 2797–2822.

Kurtz, K. J., Miao, C., & Gentner, D. (2001). Learning by analogical bootstrapping. *Journal of the Learning Sciences*, 10(4), 417–446.

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). Austin: Cognitive Science Society.

Legare, C. H. (2012). Exploring explanation: Explaining inconsistent evidence informs exploratory, hypothesis-testing behavior in young children. *Child Development*, 83(1), 173–185.

Legare, C. H. (2014). The contributions of explanation and exploration to children's scientific reasoning. *Child Development Perspectives*, 8, 101–106.

Legare, C. H., Gelman, S. A., & Wellman, H. M. (2010). Inconsistency with prior knowledge triggers children's causal explanatory reasoning. *Child Development*, 81(3), 929–944.

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

Lombrozo, T. (2012). Explanation and abductive inference. In K. J. Holyoak & R. G. Morrison (Eds.), *Oxford handbook of thinking and reasoning* (pp. 260–276). Oxford: Oxford University Press.

Lombrozo, T., & Carey, S. (2006). Functional explanation and the function of explanation. *Cognition*, 99, 167–204.

Markman, A. B. (1997). Constraints on analogical inference. *Cognitive Science*, 21(4), 373–418.

Markman, A. B., & Gentner, D. (1993). Splitting the differences: A structural alignment view of similarity. *Journal of Memory and Language*, 32, 517–535.

Markman, A. B., & Gentner, D. (1996). Commonalities and differences in similarity comparisons. *Memory & Cognition*, 24(2), 235–249.

Nokes-Malach, T. J., VanLehn, K., Belenky, D., Lichtenstein, M., & Cox, G. (2013). Coordinating principles and examples through analogy and self-explanation. *European Journal of Education of Psychology*, 28(4), 1237–1263.

Olson, D. R. (1970). *Cognitive development: The child's acquisition of diagonality*. New York: Psychology Press.

Phillips, J., Luguri, J. B., & Knobe, J. (2015). Unifying morality's influence on non-moral judgments: The relevance of alternative possibilities. *Cognition*, 145, 30–42.

Renkl, A. (2014). Toward an instructionally oriented theory of example-based learning. *Cognitive Science*, 38, 1–37.

Richey, J. E., & Nokes-Malach, T. J. (2015). Comparing four instructional techniques for promoting robust learning. *Educational Psychology Review*, 27, 181–218.

Richey, J. E., Zepeda, C. D., & Nokes-Malach, T. J. (2015). Transfer effects of prompted and self-reported analogical comparison and self-explanation. In D. C. Noelle, R. Dale, A. S. Warlaumont, J. Yoshimi, T. Matlock, C. D. Jennings, & P. P. Maglio (Eds.), *Proceedings of the 37th Annual Meeting of the Cognitive Science Society*. Austin: Cognitive Science Society.

Richland, L. E., Zur, O., & Holyoak, K. (2007). Cognitive supports for analogies in the mathematics classroom. *Science*, 316, 1128–1129.

Rittle-Johnson, B. (2006). Promoting transfer: Effects of self-explanation and direct instruction. *Child Development*, 77(1), 1–15.

Rittle-Johnson, B., & Star, J. R. (2007). Does comparing solution methods facilitate conceptual and procedural knowledge? An experimental study on learning to solve equations. *Journal of Educational Psychology*, 99(3), 561–574.

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(3), 529–544.

Rittle-Johnson, B., & Star, J. R. (2011). The power of comparison in learning and instruction: Learning outcomes supported by different types of comparisons. In J. P. Mestre & B. H. Ross (Eds.), *Psychology of learning and motivation: Cognition in education* (Vol. 55, pp. 199–225). San Diego: Elsevier.

Sagi, E., Gentner, D., & Lovett, A. (2012). What difference reveals about similarity. *Cognitive Science*, 36(6), 1019–1050.

Schwartz, D. L., Chase, C. C., Oppezzo, M. A., & Chin, D. B. (2011). Practicing versus inventing with contrasting cases: The effects of telling first on learning and transfer. *Journal of Educational Psychology*, 103(4), 759–775.

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.

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

Spellman, B. A., & Holyoak, K. J. (1996). Pragmatics in analogical mapping. *Cognitive Psychology*, 31(3), 307–346.

van Fraassen, B. C. (1980). *The scientific image*. Oxford: Oxford University Press.

Weiner, B. (1985). "Spontaneous" causal thinking. *Psychological Bulletin*, 97(1), 74–84.

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. (2013). Explanation and prior knowledge interact to guide learning. *Cognitive Psychology*, 66, 55–84.

Wong, R. M. F., Lawson, M. J., & Keeves, J. (2002). The effects of self-explanation training on students' problem solving in high-school mathematics. *Learning and Instruction*, 12(2), 233–262.

Xu, F. (2016). Preliminary thoughts on a rational constructivist approach to cognitive development: Primitives, symbols, learning, and thinking. In D. Barner & A. S. Baron (Eds.), *Core knowledge and concept change* (pp. 11–28). Oxford: Oxford University Press.