Children's Use of Causal Structure When Making Similarity Judgments

Alexandra Rett (arett@ucsd.edu)
Department of Psychology, University of California, San Diego

Jamie Amemiya (jamemiya@ucsd.edu)
Department of Psychology, University of California, San Diego

Micah Goldwater (micah.goldwater@sydney.edu.au)
School of Psychology, The University of Sydney, Sydney, Australia

Caren M. Walker (carenwalker@ucsd.edu)
Department of Psychology, University of California, San Diego

Abstract

A deep understanding of any phenomenon requires knowing how its causal elements are related to one another. Here, we examine whether children recognize similar causal structures across superficially distinct events. We presented 4- to 7-year-olds with three-variable narratives in which story events unfold according to a *causal chain* or a *common effect* structure. We then asked children to make judgments about which stories are the most similar. Results indicate that the ability to recognize and use abstract causal structure as a metric of similarity develops gradually between the ages of 4 and 7: While we find no evidence that 4-year-olds recognize the common causal structure between events, 7-year-olds have a relatively mature understanding of causal system categories when making similarity judgements. Five- and 6-year-olds show mixed success. We discuss these findings in light of children's developing causal and abstract reasoning and propose directions for future work.

Keywords: causal systems; cognitive development; abstraction; similarity judgements

Introduction

A deep understanding of any phenomenon, such as the rise of global warming or the spread of bacteria, requires understanding the abstract causal structure that underlies it. For example, to understand how an individual contracted salmonella, one must recognize a sequence of causally-related events. Specifically, the overuse of antibiotics in farm animals can lead to antibiotic resistance, causing bacteria to be present in food, which in turn leads to bacterial infections in humans. This particular sequence can be more generally described in terms of a *causal chain*, in which variable A (antibiotics) leads to variable B (antibiotic resistance), which causes variable C (bacteria in animal products), eventually leading to outcome D (infection in humans).

In the current research, we examine children's developing ability to recognize instances of the same causal structure across superficially distinct events. In particular, we explore the development of children's sensitivity to *causal system categories*[1], or abstract patterns of causation that apply across phenomena (Rottman et al., 2012). For example, the causal chain that can be used to describe the transmission of bacteria from animals to humans can also be applied to the process by which genes dictate eye color: genes determine how much melanin is produced in the eye, which in turn determines the color of the iris. Recognition of causal system categories may facilitate knowledge transfer by supporting the learner's understanding of the *type* of relationships that underlie multivariate events. For example, comprehension of the sequence linking genes to eye color may be facilitated by drawing upon one's prior understanding of the causal chain underlying antibiotic resistance. Noticing abstract causal patterns may therefore help the learner to make sense of new information.

Existing evidence suggests that both children and adults are sensitive to differences in causal structure when making inferences about the same set of variables (Gopnik et al., 2001; Lagnado & Sloman, 2002; Muentner & Bonawitz, 2017; Schulz et al., 2007). They can also apply this abstract knowledge when acting on a causal system (e.g., Bramley et al., 2015; Lapidow & Walker, 2020; Meng et al., 2018), or when considering counterfactual possibilities (e.g., Nyhout & Ganea, 2019; Schulz et al., 2007). Further, by the preschool years, children are capable of forming second-order generalizations in several domains (e.g., Dewar & Xu, 2010; Markman, 1989; Shipley, 1993; Tenenbaum et al., 2011), and can reason about higher-order relational concepts (Holyoak & Liu, 2021).

Causal systems can be classified as a special type of relational category, in which membership is determined by a particularly abstract form of common relational structure (Rottman et al., 2012). In contrast, consider the first-order relation, *carnivore*. Although the individual elements involved may change (e.g., a shark eats a seal, a lion eats an antelope), the specific relation between them remains the same. Causal systems inherently operate at a higher level of

[1] Note, in the current paper we use the terms *causal system* and *causal structure* interchangeably.

abstraction than this type of first-order relation between entities. That is, these systems capture the causal relations among the events described by first-order relations (i.e., the relations among the relations). Thus, recognition of causal system categories may be more difficult for children to grasp than other types of relations, as they require both first-order representations of the causal relations among events and a higher-order representation of how those are related to each other.

Further, although even infants are capable of forming some second-order generalizations, recognizing common structure across distinct events may pose a challenge for young children. First, early learners often have difficulty ignoring salient perceptual cues when making similarity judgements (e.g., Gentner & Ratterman, 1991). For example, to recognize that the spread of bacteria and the genetic determination of eye color both depend on a causal chain, the learner must ignore the unique surface features of each. Conversely, abstracting causal structure can also be challenging when surface-level commonalities distract from the shared structure between events. Focusing on the fact that both infections and eye color involve biological systems, for example, might lead the learner to miss content-independent, structural similarities. Indeed, although adults are able to recognize causal system categories, they often fail to spontaneously privilege these features over more salient surface cues without additional scaffolding (Cooperrider et al., 2017 ; Rottman et al., 2012; Goldwater & Gentner, 2015).

Although children's ability to abstract shared causal structure across distinct domains has not yet been examined, related work suggests that children *can* abstract higher-order information from stories or events. For example, prior to age 10, children often fail to spontaneously abstract the moral lessons from a story, instead focusing on specific story details, or defaulting to familiar moral lessons (Gentner & Toupin, 1986; Williams et al., 2002). However, when provided with a prompt to explain story events, children as young as 5 are able to abstract the theme of the story (e.g., Walker & Lombrozo, 2017). This indicates that although they may initially fail to prioritize "deep" structure, young children are *capable* of using abstract information to support learning and transfer from narratives.

Similar results have been found in the analogical reasoning literature. For example, when asked to solve a novel problem, 4- to 5-year-olds successfully applied solutions that were structurally similar, but superficially distinct (e.g., passing objects through a tube to transfer them over a barrier; Brown et al., 1986). Critically, however, children's success in this task was mediated by their ability to recall the "goal structure" (i.e., the protagonist, the goal, the obstacle, and the solution) of the source problem. Like adults, when minimal scaffolding is provided, young children generally succeed on analogical transfer tasks (e.g., Christie & Gentner, 2014; Walker & Lombrozo, 2017). Taken together, this prior work suggests that young children are already capable of reasoning about abstract relations, including themes, schemas, and relational categories.

To explore whether children are able to recognize causal system categories, we presented them with sets of three schematized stories and asked them to match the two they judged to be the most similar. The stories were governed by one of two types of causal structures: (1) a causal chain, in which variable X leads to variable Y, which in turn causes outcome Z, and (2) a common effect, in which two variables, X and Y, both lead to the same outcome, Z. Given the large body of evidence suggesting that children's ability to abstract relational information often depends on explicit scaffolding or an overall high degree of similarity between the items to be compared (e.g., Brown & Kane, 1988; Christie & Gentner, 2014), we created stories that followed a consistent narrative structure. However, we did *not* explicitly prompt children to attend to causal structure, nor did we indicate any mappings between the stories. In order to establish whether children accurately represented the causal structure embedded within the narratives, we also asked them a counterfactual question about each story (see Woodward, 2003). Finally, we did not include conflicting domain information that could provide an alternate source of similarity. In other words, rather than asking whether children *prioritize* abstract structure over surface content, we first examine whether children treat causal structure as a reasonable metric for similarity in the absence of competing cues.

As this study represents a first step in exploring children's recognition of shared causal structure across superficially distinct events, we included a relatively broad range of ages (4- to 7-year-olds), which was motivated by related work in causal and abstract reasoning. Although we did not have a specific developmental prediction, prior work examining causal reasoning can be interpreted to support two alternatives: If the ability to spontaneously recognize shared causal patterns parallels the development of reasoning about simple causal systems (e.g., Nyhout & Ganea, 2019; Lapidow & Walker, 2020; Schulz et al., 2007), it may emerge relatively early. On the other hand, there is some evidence suggesting that the ability to reason about more complex, three-variable causal systems may not appear until several years later (e.g., Frosch et al., 2012; McCormack et al., 2016; Meng et al., 2018). Work in the development of analogical reasoning (e.g., Gentner, 2010) has also proposed that this capacity may be relatively late-developing. Thus, this study serves as an initial attempt to examine whether and when children view shared causal structure as a metric of similarity between unique events.

Method

Participants

A total of sixty-four 4- to 7-year-olds participated in the study (39 female), with 16 children in each age group (*4-year-olds:* $M = 4.6$ years, $SD = .36$; *5-year-olds:* $M = 5.6$ years, $SD = .31$; *6-year-olds:* $M = 6.4$ years, $SD = .21$; *7-year-olds:* $M = 7.5$ years, $SD = .32$). An additional 11 children were tested but excluded due to parental interference (n = 3), technical difficulties (n = 4), experimenter error (n = 3), or failure to complete the study (n = 1). The study was conducted online, and participants were primarily recruited via email through a pre-existing database. Fifty-four participants provided self-identified demographic data, with the majority identifying as White (61%), multiracial (24%), or Asian (15%).

Materials

A PowerPoint presentation was shared with participants over Zoom. We created six 3-variable stories about simple events with content that is familiar to young children. The stories were designed to be presented as either a causal chain (e.g., "The sun shone very bright and warm, which made the dirt dry. The dry dirt made the flowers droopy.") or common effect (e.g., "The dirt was very dry, which made the flowers droopy. The sun also shone very bright and warm, which also made the flowers droopy.”).[2] We selected these particular causal structures for two reasons: (1) Prior work has shown that children are capable of inferring both of these structures from their observations (e.g., Schulz et al., 2007), and (2) the contrast between them is particularly salient (i.e., a causal chain originates from a *single* event, while a common effect originates from *two* events). The particular causal structure of each story was counterbalanced across participants, and the audio for each story was pre-recorded to ensure consistency.

Four of the stories were “exemplar” stories, two of which were introduced at the beginning of each trial. The remaining two stories were used as “test” stories, with one presented on each trial. Three blank boxes accompanied by arrows were used for each of the exemplar stories as a visual reminder of causal structure, with three images illustrating the events in each story later placed inside the boxes. In contrast, the two test stories were each accompanied by a single image. This was done to prevent children from using low-level visual cues when matching the test story to the exemplar stories (i.e., matching based on the spatial layout of the boxes).

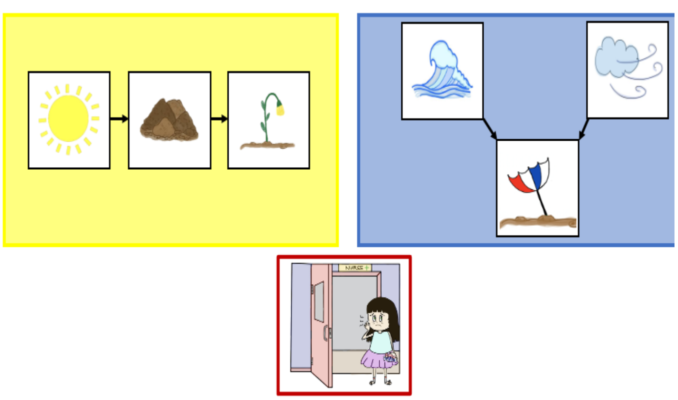

Figure 1. Still image of the participant’s screen during the test trial. The two exemplar stories appeared above the test story.

Procedure

Children were tested individually using the video conferencing platform, Zoom, with the self-view feature hidden. Prior to the study, the experimenter shared their screen and presented the game via a PowerPoint presentation.

At the start of the study, children were told they were going to listen to two different types of stories. They were shown an image of two boxes connected by an arrow (Fig. 1). The experimenter explained that for each of the stories, boxes show things that happen in the story, and arrows show how one thing makes another thing happen.

Each child then completed the two trials of the experiment. Within each trial, children listened to two “exemplar” stories. One exemplar story always described events according to a causal chain and the other story always described events according to a common effect, with the order of the stories counterbalanced between subjects.

Each story was introduced alongside a visual depiction of its causal structure (Fig. 1). The experimenter began by explaining how the different events in the story were connected to one another. For example, for the causal chain story, the experimenter said, “In this type of story, one thing happens that makes another thing happen. This middle thing then makes another thing happen.” Conversely, for the common effect story, they said, “In this type of story, one thing happens that makes another thing happen. And a different thing happens that also makes this thing happen.” The empty boxes depicting the type of causal structure (e.g., three empty boxes connected in a chain) loomed on the screen to indicate each part of the story during the narration. After listening to the description of the causal structure, children heard the story (e.g., “In this story, the sun shone very bright and warm, which made the dirt dry. The dry dirt made the flower droopy.”). As the child listened to the recording of the story, images filled the empty boxes to serve as a visual reminder of each individual event.

Children were then asked a counterfactual question to ensure that they were representing the intended causal structure. The counterfactual question always asked about the intermediate event (i.e., questions targeted variable Y for the causal chain [$X \rightarrow Y \rightarrow Z$] and common effect [$X \rightarrow Z \leftarrow Y$] stories). For example, in one causal chain story, children were asked, “In this story, if the dirt didn’t get dry, would the flower still be droopy?”. We provided feedback based on children’s answers to correct or reinforce their representation of the causal structures. For example, if children responded incorrectly to the causal chain story above, the experimenter would say, “Actually, in this story, if the dirt did not get dry, the flower would not be droopy!”

The stories then moved to the top of the screen, each with a different colored background (yellow or blue). The experimenter then introduced a third story: “Now we’re going to listen to another story! This time, I’m not going to tell you what type of story it is. I want you to try to figure out if it’s more like the [yellow] type of story or more like the

[2] For complete list of stories see: https://osf.io/sfqn4/

[blue] type of story!" A single image then appeared on the screen below the yellow and blue boxes that depicted the "test" story (Fig. 1). As there was only a single image visible to remind children of the story, the experimenter played the recording of the story twice. After children heard the test story (e.g., "In this story, Shawn ate a lot of candy, which made her teeth sore. Her sore teeth made her have to go to the school nurse's office."), they were asked to make a judgment about similarity: "Is this story more like the [blue] type of story or more like the [yellow] type of story?" Children's responses were coded as correct if they chose the exemplar story with the same causal structure as the test story (1) and incorrect otherwise (0).

To assess whether children were accurately representing the causal structures of the test stories, children were asked a second counterfactual question. While the experimenter asked the question, an image appeared on the screen that showed a red "X" through one of the events in the story to serve as a visual reminder, but did not include the depiction of the entire causal structure. This time, the experimenter only provided neutral feedback to the child's response, saying "Thank you! Let's keep playing!" Responses were scored based on the corresponding causal structure (correct = 1, incorrect = 0). The procedure was then repeated with a second set of stories.

Results

Counterfactual Performance: Exemplar Stories

Across all trials, participants were asked four counterfactual questions regarding the exemplar stories (two stories per trial). Performance on the exemplar counterfactual questions generally improved with age (*4-year-olds:* M = .53, 95% CI [.29, .78]; *5-year-olds:* M = .63, 95% CI [.39, .86]; *6-year-olds:* M = .73, 95% CI [.52, .95]; *7-year-olds:* M = .64, 95% CI [.41, .88]).[3]

To analyze whether this trend was statistically significant, we ran a Generalized Estimating Equation (GEE) model (independent correlation structure, binary logistic) predicting children's responses to the exemplar counterfactual question (correct or incorrect) as a function of the following predictors: age (mean-centered, continuous), the causal structure of each story (causal chain or common effect), order, and the age by causal structure and age by order interactions. None of the variables were significant predictors (all ps > .16). Comparing the performance of each group to chance (chance = 2), we found that 6- and 7-year-olds (p < .001 and p = .03, respectively) performed significantly better than chance, while 4- and 5-year-olds did not (ps > .09).

[3] Data available at: https://osf.io/sfqn4/

Table 1. Number of children in each age group scoring at each level (out of 2) for the similarity judgment and the counterfactual test question. The proportion of children in each group is shown in parentheses.

Number of Questions Answered Correctly	Number of Children (Proportion)	
	Similarity Judgment	CF Test Question
4-year-olds		
0	4 (.25)	1 (.06)
1	5 (.31)	12 (.75)
2	7 (.44)	3 (.19)
5-year-olds		
0	2 (.13)	3 (.19)
1	5 (.31)	11 (.69)
2	9 (.56)	2 (.12)
6-year-olds		
0	2 (.13)	1 (.06)
1	6 (.37)	5 (.31)
2	8 (.50)	10 (.63)
7-year-olds		
0	0 (0)	0 (0)
1	3 (.19)	5 (.31)
2	13 (.81)	11 (.69)

Counterfactual Performance: Test Stories

Next, we examined children's responses to the counterfactual questions asked at test. Performance on the counterfactual test questions also improved with age (*4-year-olds:* M = .56, 95% CI [.32, .81]; *5-year-olds:* M = .47, 95% CI [.22, .71]; *6-year-olds:* M = .78, 95% CI [.58, .98]; *7-year-olds:* M = .84, 95% CI [.67, 1.0]).

To analyze children's performance, we again ran a GEE model predicting children's responses to the counterfactual test question (correct or incorrect) as a function of the same predictors used above. We found a significant main effect of age, B = 1.28, SE = .38, Wald χ^2 (1) = 9.81, p = .002, odds ratio = 3.61, 95% CI [1.73, 7.55]. There was also a significant main effect of causal structure, B = -1.61, SE = .45, Wald χ^2 (1) = 8.98, p = .003, odds ratio = .20, 95% CI [.08, .48], with children performing better in response to the counterfactual question for the causal chain stories (M = .90, 95% CI [.68, .92]) in comparison to the common effect stories (M = .53, 95% CI [.42, .63]). There was no main effect of order (i.e., p = .862), and no significant interaction between age and order (p = .174). The interaction between age and causal structure produced a marginally significant effect (p = .054).

To analyze the performance of each age group against chance, we compared the proportion of children who scored 2 out of 2 on the counterfactual test questions to a chance distribution of .25 using binomial tests (see Table 1 for the number of children at each score level). We chose this method as it offers a more conservative and sensitive account of children's performance compared to examining the average score for each group. The number of 6- and 7-year-olds with a score of 2/2 on the counterfactual test question was significantly higher than chance (p = .002 and $p < .001$, respectively). However, 4- and 5-year-olds responded at chance values (*ps* > .4).

Although the interaction between age and causal structure produced only a marginally significant effect, we also conducted exploratory analyses to further examine the finding that children performed relatively better on counterfactual test questions for the causal chain stories than the common effect stories. We found that 6- and 7-year-olds performed above chance on counterfactual questions for *both* the common effect and causal chain stories (*ps* < .002), indicating that success at these ages was not solely driven by ceiling performance on the causal chain. In contrast, 4-year-olds were at chance for questions about both types of causal structure, and 5-year-olds were only different from chance for questions about the causal chain stories ($p < .001$).

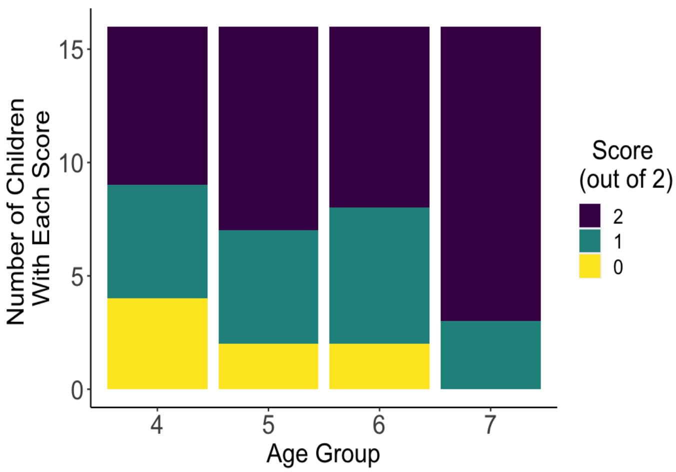

Figure 2. Number of children responding to the similarity judgment question at each score level (out of 2) by age group.

Similarity Judgments

Our main analysis of interest was whether children matched stories based on their underlying causal structure. We again ran a GEE model (independent correlation structure, binary logistic) predicting children's similarity judgments as a function of the following predictors: age (mean-centered, continuous variable), counterfactual score for the test story (correct or incorrect), order, causal structure, and the interaction between age and each predictor.

We found a significant main effect of age, B = -.254, SE = .53, Wald χ^2 (1) = 6.12, p = .013, odds ratio = 1.29, 95% CI [.45, 3.63], such that performance on the similarity judgment question improved with age (*4-year-olds: M* = .59, 95% CI [.35, .84]; *5-year-olds: M* = .72, 95% CI [.50, .94]; *6-year-olds: M* = .69, 95% CI [.46, .92]; *7-year-olds: M* = .91, 95% CI [.76, 1.0]). There were no main effects of order, counterfactual score, or causal structure (*ps* > .11), and the interactions between age and counterfactual score (p = .407), age and causal structure (p = .086), and age and order (p = .288) were also not significant.

We again analyzed the performance of each age group by comparing the number of children who scored 2/2 against a chance level of .25 using binomial tests (Fig. 2). The number of 5-, 6-, and 7-year-olds with a score of 2/2 was significantly higher than chance (p = .007, p = .040, $p < .001$, respectively). However, 4-year-olds' performance was not significantly different from chance (p = .09).

Discussion

The ability to recognize the common causal structure between events may play a central role in knowledge transfer, allowing learners to make sense of information about novel causal relationships by relating these events to their existing knowledge of causal system categories. We find that children are capable of abstracting causal system categories and using them as the basis for similarity judgements. Results suggest that this ability develops steadily between the ages of 4 to 7. While 4-year-olds responses could not be distinguished from chance, 5- and 6-year-olds begin to recognize the common causal structure between events, with mixed success. By age 7, children consistently matched events according to their common causal structure.

This trajectory is consistent with recent research in the development of causal reasoning. While even preschool-aged children represent their causal knowledge in terms of its underlying causal structure (Muentner & Bonawitz, 2017), the ability to successfully reason over these structures sometimes develops later for more challenging cases. For example, while some work demonstrates that 4- to 6-year-olds can discriminate possible causal structures when presented with a forced choice between two actions (e.g., Lapidow & Walker, 2020), other work suggests that children under 7 struggle to select interventions that disambiguate more complex three-variable systems (e.g., Frosch et al., 2012; McCormack et al., 2016; Meng et al., 2018). Relatedly, although 4- to 5-year-old children can reason counterfactually about physical events in a non-verbal paradigm (e.g., Nyhout & Ganea, 2019), this capacity continues to mature in tasks that rely on interpreting and responding to more complex narratives (e.g., Nyhout et al., 2019; Rafetseder et al., 2013). Thus, while preschool-aged children can represent causal system categories, their ability to use this information to recognize events with a common underlying structure may continue to develop beyond the preschool years.

However, there are other potential explanations for the developmental trend we observed in the current study. First, it is relatively challenging to reason about causal structures embedded within narratives. To avoid the possibility of low-level perceptual matching between stories, children were required to recall the structure of the test stories without detailed visual aids, which may have introduced working memory demands. Second, since the narratives were based on familiar, everyday scenarios, this may have led to interference from children's prior knowledge. That is, children may have spontaneously inferred relationships between variables, based on their own experience. In some cases, these inferences may have conflicted with the causal structures that were presented in the task. To address these issues, future work will replace these narratives with three-variable physical systems (e.g., simple machines) to reduce some of these task demands.

One surprising result from the current study was that correctly responding to the counterfactual test question did not predict similarity judgments (though 7-year-olds performed above chance on the counterfactual test question and near ceiling on similarity judgements). It is possible that younger children may have defaulted to a simple strategy for the counterfactual test question, due to working memory demands. That is, some children may have only considered the relationship between the two variables mentioned in the counterfactual question (e.g., whether a change to the dirt would impact the flowers, regardless of the third variable), leading to an increase in “no" responses. This strategy would lead children to provide the correct response for the causal chain, but not for the common effect. In line with this possibility, children performed better on the counterfactual test question for the causal chain stories. However, children’s similarity judgments did not depend on the test story’s causal structure, making it unlikely that children simply understood the causal chain stories better. Indeed, further, exploratory analyses indicated that 6- and 7-year-olds performance on counterfactual test questions was not solely driven by ceiling performance on questions about the causal chain stories. Despite this, given the potential confound introduced by a potential “no” bias, additional measures of causal understanding will be included in future work.

One possible alternative explanation for why performance on counterfactual questions did not predict the accuracy of similarity judgments is that robust counterfactual reasoning may not be necessary for recognizing common causal structure. Instead, it is possible that these two forms of reasoning may emerge in parallel. Further, asking children to compare stories may have facilitated their representation of the relevant causal structures, boosting performance for both counterfactual questions and similarity judgments. In line with this possibility, related work on analogical reasoning finds that the process of comparison itself can support children’s recognition of key relations (Gentner & Markman, 1997). Additional work is needed to further examine the relationship between counterfactual reasoning and the recognition of causal system categories.

A possible limitation of the current study is that, although children were unable to rely on low-level *visual* cues when matching stories, similarity judgements may have been influenced by shared syntax across structurally similar narratives. To simplify the study design and facilitate abstraction, we intentionally used consistent language across stories (e.g., “The dirt was very dry, which made the flowers droopy. The sun also shone very bright and warm, which also made the flowers droopy.”). If some children matched based on similarity of causal language, they may have succeeded without abstracting shared causal structure. On the other hand, it is reasonable to assume that children’s inferences about causality are partly scaffolded by attention to these linguistic cues, particularly when causal relationships are presented in a narrative context. In fact, children may initially become aware of causal system categories by recognizing consistencies in the way that these events are described (Muentener & Schulz, 2012). If so, this would parallel work suggesting that relational language (e.g., *on, in,* or *under* to denote spatial relations) can foster abstract reasoning (e.g., Gentner, 2016). To further explore the possibility that causal language supports children’s similarity judgements, future studies will control for these syntactic cues.

Here, we present preliminary evidence that children are capable of abstracting causal system categories. This result adds to the literature on children’s causal reasoning, and opens up new avenues for applied work exploring how the recognition of causal systems may support learning and transfer in children.

Acknowledgements

Thank you to Brendan Hwang, Mia Real, Nazgol Behdinan, and Trisha Katz for their efforts towards data collection and recruitment. We also thank all of the participating families who made this research possible. This work was supported by an NSERC PGS-D to A. Rett, the Eunice Kennedy Shriver National Institute of Child Health & Human Development of the NIH (Award Number: F32HD098777) to J. Amemiya, and the Jacobs Foundation Fellowship to C. Walker.

References

Bramley, N. R., Lagnado, D. A., & Speekenbrink, M. (2015). Conservative forgetful scholars: How people learn causal structure through sequences of interventions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(3), 708.

Brown, A. L., & Kane, M. J. (1988). Preschool children can learn to transfer: Learning to learn and learning from example. *Cognitive Psychology*, *20*(4), 493-523.

Brown, A. L., Kane, M. J., & Echols, C. H. (1986). Young children’s mental models determine analogical transfer across problems with a common goal structure. *Cognitive Development*, *1*, 103–121.

Christie, S., & Gentner, D. (2014). Language helps children succeed on a classic analogy task. *Cognitive Science*, *38*, 383–397.

Cooperrider, K., Gentner, D. & Goldin-Meadow, S. (2017). Analogical gestures foster understanding of causal systems. *Proceedings of the 39th Annual Meeting of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

Dewar, K. M., & Xu, F. (2010). Induction, overhypothesis, and the origin of abstract knowledge: Evidence from 9-month-old infants. *Psychological science*, *21*(12), 1871-1877.

Frosch, C. A., McCormack, T., Lagnado, D. A., & Burns, P. (2012). Are causal structure and intervention judgments inextricably linked? A developmental study. *Cognitive Science*, *36*(2), 261-285.

Gentner, D. (2016). Language as cognitive tool kit: How language supports relational thought. *American Psychologist*. 71(8):650-657.

Gentner, D., & Markman, A. B. (1997). Structure mapping in analogy and similarity. *American psychologist*, *52*(1), 45.

Gentner, D. & Rattermann, M. J. (1991) Language and the career of similarity. In: *Perspectives on language and thought: Interrelations in development*, ed. S. A. Gelman & J. P. Byrnes, pp. 225–77. Cambridge University Press.

Gentner, D., & Toupin, C. (1986). Systematicity and surface similarity in the development of analogy. *Cognitive science*, *10*(3), 277-300.

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

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.

Gopnik, A., Sobel, D. M., Schulz, L. E., & Glymour, C. (2001). Causal learning mechanisms in very young children: two-, three-, and four-year-olds infer causal relations from patterns of variation and covariation. Developmental psychology, 37(5), 620.

Holyoak, K. J., & Lu, H. (2021). Emergence of relational reasoning. *Current Opinion in Behavioral Sciences*, *37*, 118-124.

Lagnado, D. A., & Sloman, S. (2002). Learning causal structure. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 24, No. 24).

Lapidow, E., & Walker, C. M. (2020). Informative experimentation in intuitive science: Children select and learn from their own causal interventions. Cognition, 201, 104315.

Markman, E. M. (1989). *Categorization and naming in children: Problems of induction*. mit Press.

McCormack, T., Bramley, N., Frosch, C., Patrick, F., & Lagnado, D. (2016). Children's use of interventions to learn causal structure. *Journal of experimental child psychology*, 141, 1-22.

Meng, Y., Bramley, N., & Xu, F. (2018). Children's causal interventions combine discrimination and confirmation. In *Proceedings of the 40th annual conference of the Cognitive Science Society.*

Muentener, P., & Bonawitz, E. B. (2017). The development of causal reasoning. In M. R. Waldmann (Ed.), *The Oxford handbook of causal reasoning* (pp. 677–698). New York: Oxford University Press.

Muentener, P., & Schulz, L. (2012). What doesn't go without saying: Communication, induction, and exploration. *Language Learning and Development*, *8*(1), 61-85.

Nyhout, A., & Ganea, P. A. (2019). Mature counterfactual reasoning in 4-and 5-year-olds. *Cognition*, 183, 57-66.

Nyhout, A., Henke, L., & Ganea, P. A. (2019). Children's counterfactual reasoning about causally overdetermined events. Child development, 90(2), 610-622.

Rafetseder, E., & Perner, J. (2010). Is reasoning from counterfactual antecedents evidence for counterfactual reasoning?. *Thinking & reasoning*, *16*(2), 131-155.

Rafetseder, E., Schwitalla, M., & Perner, J. (2013). Counterfactual reasoning: From childhood to adulthood. *Journal of experimental child psychology*, *114*(3), 389-404.

Rottman, B. M., Gentner, D., & Goldwater, M. B. (2012). Causal systems categories: Differences in novice and expert categorization of causal phenomena. *Cognitive science*, 36(5), 919-932.

Schulz, L. E., Gopnik, A., & Glymour, C. (2007). Preschool children learn about causal structure from conditional interventions. *Developmental science*, 10(3), 322-332.

Shipley, E. F. (1993). Categories, hierarchies, and induction. *The psychology of learning and motivation*, *30*, 265-301.

Tenenbaum, J. B., Kemp, C., Griffiths, T. L., & Goodman, N. D. (2011). How to grow a mind: Statistics, structure, and abstraction. *science*, *331*(6022), 1279-1285.

Walker, C. M., & Lombrozo, T. (2017). Explaining the moral of the story. *Cognition*, *167*, 266-281.

Williams, J. P., Lauer, K. D., Hall, K. M., Lord, K. M., Gugga, S. S., Bak, S. J., ... & deCani, J. S. (2002). Teaching elementary school students to identify story themes. *Journal of Educational Psychology*, *94*(2), 235.

Williams, J. J., & Lombrozo, T. (2010). The role of explanation in discovery and generalization: Evidence from category learning. Cognitive science, 34(5), 776-806.

Woodward, J. (2003). Making things happen: A theory of causal explanation. Oxford University Press.