Journal of Experimental Child Psychology 202 (2021) 105008

Contents lists available at ScienceDirect

Journal of Experimental Child Psychology

journal homepage: www.elsevier.com/locate/jecp

Brief Report

When correlation equals causation: A behavioral and computational account of second-order correlation learning in children

Deon T. Benton [a,*], David H. Rakison [b], David M. Sobel [a]

[a] Department of Cognitive, Linguistics, & Psychological Sciences, Brown University, Providence, RI 02912, USA
[b] Department of Psychology, Carnegie Mellon University, Pittsburgh, PA 15213, USA

ARTICLE INFO

Article history:
Received 25 February 2020
Revised 11 September 2020
Available online 20 October 2020

Keywords:
Causal learning
Computational modeling
Second-order correlation learning
Causal reasoning
Causal representations
Neural networks

ABSTRACT

We examined 2- and 3-year-old children's ability to use second-order correlation learning—in which a learned correlation between two pairs of features (e.g., A and B, A and C) is generalized to the noncontiguous features (i.e., B and C)—to make causal inferences. Previous findings showed that 20- and 26-month-old children can use second-order correlation learning to learn about static and dynamic features in category and noncategory contexts. The current behavioral study and computational model extend these findings to show that 2- and 3-year-olds can detect the second-order correlation between an object's surface feature and its capacity to activate a novel machine, but only if the children had encoded the first-order correlations on which the second-order correlation was based. These results have implications for children's developing information-processing capacities on their ability to use second-order correlations to infer causal relations in the world.

Published by Elsevier Inc.

* Corresponding author.
E-mail address: dbenton2@swarthmore.edu (D.T. Benton).

https://doi.org/10.1016/j.jecp.2020.105008
0022-0965/Published by Elsevier Inc.

Introduction

Causal reasoning is a fundamental part of child development. Children must learn to encode causal structure to make predictions, generate explanations, and reason counterfactually. For example, inferring that flipping a light switch to *on* will cause a light bulb to illuminate requires recognizing that light switches cause light bulbs to turn on and not the other way around and that light bulbs generally do not illuminate when their switches are *off*.

There is now considerable evidence showing that infants possess statistical learning mechanisms that support learning across a variety of domains (e.g., Haith, 1993; Kirkham, Slemmer, & Johnson, 2002; Saffran, Aslin, & Newport, 1996). For example, infants between 4½ and 10 months of age can use statistical information as well as spatial and temporal cues to perceive causality (Cohen & Amsel, 1998; Leslie & Keeble, 1987; Oakes & Cohen, 1990; Rakison & Krogh, 2012). By 2 years of age, children make causal inferences that extend beyond the statistical regularities that they observe (e.g., Luchkina, Sommerville, & Sobel, 2018; Meltzoff, Waismeyer, & Gopnik, 2012; Sobel & Kirkham, 2006; Waismeyer & Meltzoff, 2017; Waismeyer, Meltzoff, & Gopnik, 2015; Walker & Gopnik, 2014; Walker, Lombrozo, Williams, Rafferty, & Gopnik, 2017). These findings suggest that infants might initially perceive and encode statistical regularity among events, but how these capacities support causal reasoning in children is unclear.

One promising mechanism that may assist in making causal inferences from statistical regularity is *second-order correlational learning*. This refers to the ability to make inferences about features that are correlated but are separated in time and space. For example, it is commonplace for infants to register that hands (A) convey information about goal-directed action (B) (e.g., Woodward, 1998). But objects with hands (A) often also have legs (C). Can infants infer that objects with legs are goal-directed? That is, given a correlation between A and B and between A and C, do infants also register that B and C should co-occur? Second-order correlation is a powerful domain-general learning process because learners need not be exposed to the full space of correlations for inferences to be made. For example, a learner who is presented with the two sets of aforementioned relations (i.e., A and B and A and C) may infer—through second-order correlation learning—that a novel object with legs is goal-directed.

Several studies have demonstrated that young children can detect second-order correlations among the static features of objects (Cuevas, Rovee-Collier, & Learmonth, 2006; Yermolayeva & Rakison, 2016). Rakison and Benton (2019) showed that toddlers also make such inferences about the dynamic features of objects. They habituated 20- and 26-month-olds to separate correlations between features of objects. First, they presented toddlers with stationary objects of different shapes that had different surface features (e.g., blue squares had yellow hearts inside; red circles had white crosses inside). The second involved a dynamic relation between an object and its motion trajectory (featureless blue squares and blue pentagons moved rectilinearly; featureless red circles and triangles moved curvilinearly). The study tested whether toddlers extracted the second-order correlation between yellow hearts and rectilinear motion and between white crosses and curvilinear motion.

Toddlers were then shown a test event that was consistent with these second-order correlations (a novel object with a yellow heart inside moved rectilinearly) and a test event that was inconsistent with these correlations (a novel object with a yellow heart inside moved curvilinearly). Both age groups detected the second-order correlation: the 20-month-olds looked longer at the inconsistent test event, whereas the 26-month-olds looked longer at the consistent test event. Given that the features were presented equally during habituation, greater interest in either test event could occur only if infants learned the second-order correlation between features and motion paths. Rakison and Benton (2019) also showed that a computational (connectionist) model provided a good description of children's inferences, including the developmental difference in looking times. These results indicated that children (and the computational model) could detect the correlation between features that are not presented together simultaneously.

The goal of the current investigation was to demonstrate that toddlers make similar inferences about the causal properties of objects. We examined whether 2- and 3-year-olds could detect a second-order correlation between objects' surface features and their causal efficacy. Critically, examining this age group allowed us to expand on the Rakison and Benton (2019) findings and to address an important facet of children's causal reasoning. Numerous studies on toddlers' causal reasoning suggest that various information-processing demands influence children's ability to make certain inferences (e.g., Luchkina et al., 2018; Walker & Gopnik, 2014). To detect a second-order correlation, toddlers must be able to detect the first-order correlations that form the basis of the inference. We examined whether children's inferences are influenced by this capacity in a behavioral study. We then presented a computational model similar to that of Rakison and Benton (2019). The computational model enabled us to determine to what extent second-order correlation learning for causal stimuli emerges in an associative-learning system that embodies certain information-processing constraints such as constraints on the speed of learning and on the amount of information that is retained across the phases of the experiment.

Behavioral study

In the behavioral study, 2- and 3-year-olds were shown two objects, each with a distinct correlated feature (e.g., Object 1 with Feature A and Object 2 with Feature B). Children then observed that Object 1 without Feature A activated a novel machine, whereas Object 2 did not. Children were then asked to choose which of two new objects—one with Feature A and one with Feature B—would activate the machine. We predicted that children would choose the test object with Feature A if they detected the indirect relation between that feature and the machine's activation but that this choice would depend critically on their encoding of the first- and second-order associations.

Method

Participants

Participants were 32 2-year-olds (17 boys and 15 girls; M_{age} = 29.21 months, range = 24–35) and 32 3-year-olds (17 boys and 15 girls; M_{age} = 39.83 months, range = 36–47). Sample size for all analyses was determined based on a power analysis for χ^2 tests (α = .05, power = .80, effect size = .50). An additional 3 children were tested but not included in the final analysis because of experimenter error (n = 2) or because the child refused to participate in the experiment (n = 1). All participants were tested in a quiet room in a local children's museum. No explicit information about ethnicity or socioeconomic status was collected, but previous studies at the museum suggest the following breakdown: 14% Hispanic, 3% African American, 5% Asian American, 1% Native American, 58% Caucasian, 7% mixed race, and 12% no response.

Materials

This experiment used a version of the *blicket detector* (Gopnik & Sobel, 2000). The machine was a 12.7 × 17.8 × 7.6-cm black plastic box with a translucent white top. Pressing a button on a wireless remote control activated the machine, and the experimenter activated the machine as soon as an object contacted it. This created the illusion that objects activated the machine. Six 3.8 × 2.54-cm wooden blocks and two unique stickers were also used (Fig. 1).

Procedure

Children were tested in a quiet room in a local museum and were seated across the table facing a male experimenter. The experiment consisted of two training trials and a test phase.

Training trials. In the static training trial, children were first shown a red cube and a green cylinder, which were positioned next to each other on the table. One object had a yellow circle sticker on it (e.g., the green cylinder), and the other object had a purple diamond sticker on it (e.g., the red cube). The correlation between the blocks and stickers was counterbalanced across participants. Children were shown each block one at a time and were told, "See this toy? This toy has a yellow [or red] thing on it," This was done twice. The order of presentation was random.

In the causal training trial, two new objects were placed on the table. These were sticker-less objects that were identical to the first pair of objects. The experimenter then introduced the machine.

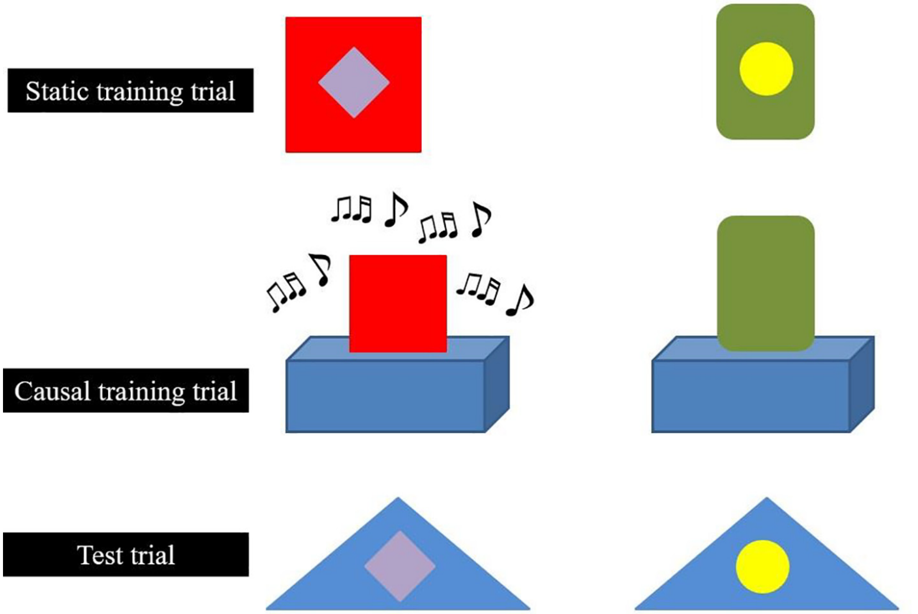

Fig. 1. Examples of the stimuli used in the study. Children were shown the static training trial first, followed by the causal training trial. Two novel, identical objects were presented at test. The inconsistent test object had the surface feature that was indirectly associated with the machine's nonactivation. The consistent test object had the surface feature was that indirectly associated with the machine's activation. Children were asked which object made the machine go.

Children were told that the machine "lights up and plays music when some toys are put on it, but not when other toys are put on it." The two blocks were then placed on the machine one at a time. One block made the machine go (e.g., the red cube), and the other block did not (e.g., the green cylinder); this was counterbalanced across participants.

Test trial. Two new, identical blue triangular blocks were brought out. One had the same sticker that was paired with the red cube in the static training trial, and the other new block had the same sticker that was paired with the green cylinder on the static training trial. The experimenter then said, "Here are two new toys. Can you tell me which toy makes the machine go?" If children inferred an association between the red cube and the yellow sticker during the static phase, and then an association between the red cube (without the sticker) and the machine's activation during the causal phase, then at test they should use the blue block with the yellow sticker to make the machine go and not the other block. This is because the yellow sticker was associated—via a second-order correlation—with the machine's activation (see Fig. 1).

All children were then given a memory check in which they were shown the static phase blocks and stickers and were asked to place the stickers on the blocks to which they were initially appended.

Results

All p values were supplemented with Bayes factors (*BF*s) to quantify the relative support for the alternative hypothesis compared with the null hypothesis. *BF*s close to 1 indicate equal support for both hypotheses, whereas *BF*s ≥ 3 and < 10 indicate moderate and strong evidence for the alternative hypothesis, respectively (Lee & Wagenmakers, 2014). Table 1 shows the frequency with which children passed the memory control and chose the object consistent with detecting the second-order correlation between the two age groups. Preliminary analyses showed that neither children's sex nor any

Table 1
Frequency of children choosing the test object consistent with registering the second-order correlation.

	2-year-olds	3-year-olds
Responded correctly on memory question (n = 47)	10 of 17	22 of 30
Responded incorrectly on memory question (n = 17)	9 of 15	1 of 2
Total	19 of 32	23 of 32

of our counterbalancing measures was significantly related to their choice of the test object, all $\chi^2(1, N = 64)$ values < 1.91, all p values > .16.

We next examined the relation between children's age and their choices on the test question and memory control. All 64 participants were used in these analyses. Although there was no relation between children's age (in months) and their choice of object at test, $r_s(62) = .02$, $p = .910$, there was a significant relation between children's age and their performance on the memory control, $r_s(62) = .40$, $p = .001$. A logistic regression corroborated these results. This analysis revealed that age did not significantly predict choices on the test question, odds ratio = 1.01, $p = .98$, 95% confidence interval (CI) [−1.41, 3.44], $BF = 0.13$, but did significantly predict performance on the memory control, odds ratio = 7.13, $p < .01$, 95% CI [0.11, 14.14], $BF = 25.00$. Children who passed the memory control were more likely to choose the correct test object than children who did not pass it, $\chi^2(1, N = 64) = 6.15$, $p = .01$, $BF = 5.37$. Follow-up comparisons indicated that children who passed the memory control chose the consistent test object 68% of the time (32 of 47 children), which is greater than expected by chance, binomial test, $p = .02$, $BF = 5.37$. In contrast, children who did not pass the memory control chose the consistent test object 59% of the time (10 of 17), which is not significantly different from chance, binomial test, $p = .63$, $BF = 0.62$. This analysis suggests that 2- and 3-year-olds do not differ in their ability to use second-order correlational information but potentially differ in their ability to encode the first-order correlations on which the second-order correlation is built.

Discussion

Both 2- and 3-year-olds registered second-order correlations to make causal inferences, but only if they remembered the feature–block pairings from the static phase. The ability to remember the initial correlation in this procedure develops at 2 or 3 years of age. This result is consistent with other findings suggesting that even younger children may be capable of encoding these correlations (Cuevas et al., 2006; Yermolayeva & Rakison, 2016). In those studies, the initial association between the stimuli was presented for a much longer time, which may have allowed for better encoding of the information.

We have suggested that second-order correlational learning is a domain-general associative-learning mechanism that requires sufficient information-processing capacities. To examine this hypothesis, we constructed a similar computational model (i.e., an autoencoder connectionist model) to that of Rakison and Benton (2019) to assess whether it can explain children's second-order correlational learning and performance on the memory control. The success of the model in simulating children's second-order correlation learning performance as well as their performance on the memory control is important because it would suggest that an associative-learning mechanism with sufficient information-processing capacities can form second-order correlations for causal stimuli.

Computational model

Method

Network architecture

We used a three-layer, autoencoder neural network that was trained using backpropagation and momentum. The learning rate, momentum, weight decay, and number of hidden units were set to .08, .90, .001, and 15, respectively, and corresponded to children who passed the memory check and were set to .001, .90, .005, and 13, respectively, and corresponded to children who failed the memory check. This implemented a simple model of development capable of simulating the difference in choices on the test question between children who passed the memory check and those who failed it. A total of 32 networks—each initialized with small (distribution range = ±0.8) random weights—were run for each simulation. This number paralleled the number of participants in the behavioral study. Finally, both network types received 50 training epochs (or trials) because children who passed and those who failed the memory check in the behavioral study received an equal number of training trials. Our number of training epochs, 50, was chosen to mirror Rakison and Benton (2019), who used the number of training epochs as a way of simulating the "age" of the model. A total of 50 training epochs simulated the models for the 26-month-olds (which would be akin to the youngest participants tested here). Nonetheless, the simulation results reported below held regardless of whether models were trained for as few as 15 to 20 epochs or as many as 200 to 250 epochs.

The input to the network consisted of patterns of activity across three different groups of input units: shape, color, and feature input groups (see Fig. 2). The shape group consisted of distributed activity across 12 units, whereas activity was encoded locally across 3 units for the color group and across 2 units for the feature group. Critically, the similarity between any two input patterns within each of the three groups was orthogonal to ensure unbiased output responses. Each input group projected to a group of hidden units, which in turn were connected to output groups that were copies of each input group and an additional single-unit output group that took on a value of 1 whenever a causal object was presented and a value of 0 otherwise.

We represented the task in terms of three input groups for three reasons. First, these dimensions would be available to low-level perceptual processes that segment tasks and stimuli into component parts. Second, representing the task and stimuli in terms of three feature groups would be the absolute

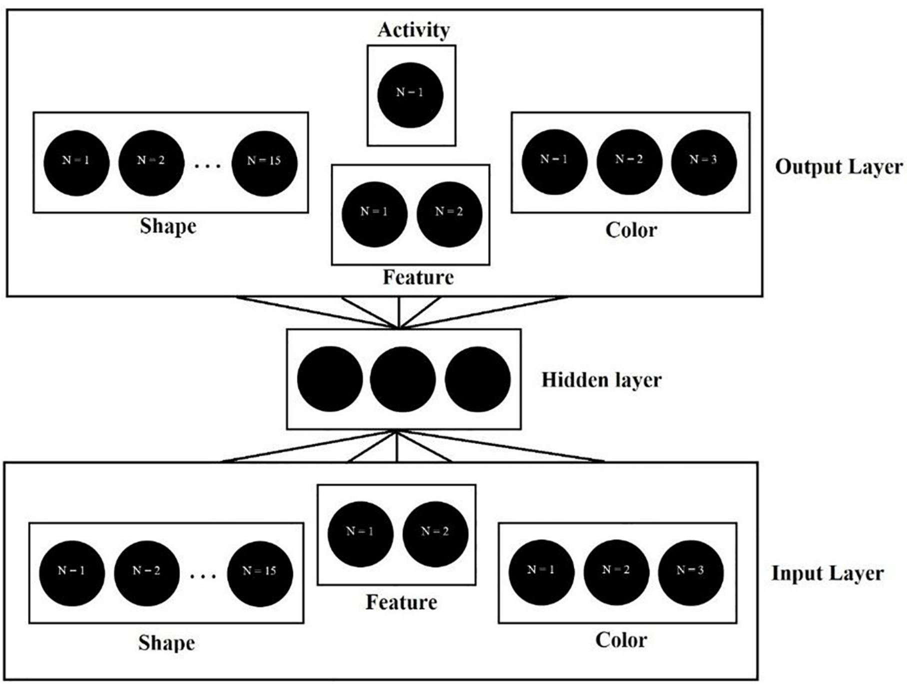

Fig. 2. Schematic of the model that was used to simulate the results from the behavioral study. The three input groups were connected to one hidden layer, which in turn was connected to three output groups that mirrored the three input groups. The task was to recreate the input patterns along the three input groups on each of the corresponding output groups as well as to respond with a 0 or 1 in the activity group if a causally inefficacious or efficacious object was presented, respectively.

minimum number of input groups needed to simulate successful second-order correlation learning. Third, the simulations were designed to account for children's second-order correlation learning and memory control performance rather than to capture the precise stimuli that children attended to during training and test. We do not suggest that these simulations are full cognitive models of second-order correlation learning in children. Rather, they serve as proof of concept that these models can register second-order correlations from only the statistical regularity in the data.

Training

In total, 32 simulations were run to simulate the successful networks and 32 for the unsuccessful networks. Individual networks were simulated by initializing a fresh set of small, random weights. During the static training phase, networks were trained to predict an internal circle shape on the output when given a stationary red object as input and to predict an internal diamond shape on the output when given a stationary green object. The feature–block pairings presented to networks during the static training phase were counterbalanced. During the causal training phase, networks were trained to predict either that a featureless red cube or a green cylinder made the machine go (counterbalanced).

Testing

Networks were then presented with two identically shaped objects. One of the test objects possessed the feature that was indirectly associated with the machine's activation, and the other object possessed the feature that was indirectly associated with the machine's inactivation. Correct test performance was indicated by activity in the single activity output group that was above 0.5 for the object that was designated as the correct test object, whereas chance or incorrect performance was indicated by equal activity in the activity output unit for both test objects.

Results

All of the networks with parameters that were more consistent with passing the memory control placed greater weight on the consistent test object (32 of 32 networks), binomial test, $p < .001$, $BF = 9105.00$. Networks with parameters that were more consistent with not passing the memory control were at chance in their test object choice (16 of 32), binomial test, $p = 1.00$, $BF = 0.40$. Critically, the number of consistent choices on the test question for networks that passed the memory control ($n = 32$) did not differ from that for children who passed the memory control ($n = 47$), $\chi^2(1, N = 79) = 0$, $p = 1.00$, $BF = 1.01$. Likewise, networks that did not pass the memory control ($n = 32$) did not differ from children who did not pass it in their choice of the consistent object at test ($n = 17$), $\chi^2(1, N = 49) = 1.38$, $p = .24$, $BF = 0.76$.

To ensure that these results were not idiosyncratic to the precise parameter values used in the reported simulation, we ran 384 additional simulations with modified parameter values. For networks that simulated children who passed the memory control (Table 2), the number of hidden units varied from 5 to 15 ($N = 48$), weight decay varied from 0 to .003 ($N = 48$), momentum varied from .875 to .900 ($N = 48$), and learning rate varied from .004 to .006 ($N = 48$). For networks that simulated children who failed the memory control (Table 3), the number of hidden units varied from 11 to 13 ($N = 48$), weight decay varied from .005 to .006 ($N = 48$), momentum varied from .875 to .900 ($N = 48$), and learning rate varied from .0008 to .001 ($N = 48$). As can be seen in Tables 2 and 3, models that simulated the performance of children who passed the memory control were more likely to choose the test object that embodied the consistent second-order correlation relation than the one that embodied the inconsistent relation. In contrast, models that simulated the performance of children who did not pass the memory control chose equally between test objects that embodied the consistent second-order correlation relation and those that did not. Crucially, these additional simulations replicate the original model results.

Table 2
Effects of changing parameters on the computational model.

Networks corresponding to children who passed the memory control					
Learning rate = .006		Learning rate = .005		Learning rate = .004	
Binomial	*BF*	Binomial	*BF*	Binomial	*BF*
(16 of 16), $p < .001$	585	(14 of 16), $p < .01$	18	(15 of 16), $p < .001$	83
Hidden units = 15		Hidden units = 10		Hidden units = 5	
Binomial	*BF*	Binomial	*BF*	Binomial	*BF*
(16 of 16), $p < .001$	585	(16 of 16), $p < .001$	585	(16 of 16), $p < .001$	585
Weight decay = 0		Weight decay = .0015		Weight decay = .003	
Binomial	*BF*	Binomial	*BF*	Binomial	*BF*
(16 of 16), $p < .001$	585	(16 of 16), $p < .001$	585	(15 of 16), $p < .001$	83
Momentum = .900		Momentum = .885		Momentum = .875	
Binomial	*BF*	Binomial	*BF*	Binomial	*BF*
(16 of 16), $p < .001$	585	(16 of 16), $p < .001$	585	(16 of 16), $p < .001$	585

Note. Numbers in parentheses correspond to the proportion of networks for each parameter value change that responded with the consistent object at test to the total number of networks that responded with both types of objects.

Table 3
Effects of changing parameters on the computational model.

Networks corresponding to children who did not pass the memory control					
Learning rate = .001		Learning rate = .0009		Learning rate = .0008	
Binomial	*BF*	Binomial	*BF*	Binomial	*BF*
(5 of 16), $p = .21$	1.20	(6 of 16), $p = .45$	0.75	(9 of 16), $p = .80$	0.56
Hidden units = 13		Hidden units = 12		Hidden units = 11	
Binomial	*BF*	Binomial	*BF*	Binomial	*BF*
(9 of 16), $p = .80$	0.56	(6 of 16), $p < .001$	0.75	(8 of 16), $p = 1.00$	0.52
Weight decay = .005		Weight decay = .0055		Weight decay = .006	
Binomial	*BF*	Binomial	*BF*	Binomial	*BF*
(6 of 16), $p < .001$	0.75	(5 of 16), $p = .21$	1.20	(7 of 16), $p = .80$	0.57
Momentum = .900		Momentum = .885		Momentum = .875	
Binomial	*BF*	Binomial	*BF*	Binomial	*BF*
(9 of 16), $p = .80$	0.56	(8 of 16), $p = 1.00$	0.52	(9 of 16), $p = .80$	0.56

Note. The numbers in parentheses correspond to the proportion of networks for each parameter value change that responded with the consistent object at test to the total number of networks that responded with both types of objects.

Discussion

The simulations showed that networks that possessed more robust memory capacities weighted the consistent test object more heavily than networks that possessed less robust memory capacities. Critically, this result mirrored that in the behavioral study. The improved performance of networks that corresponded to children who passed the memory control relative to those who failed it resulted from their increased processing speed (e.g., higher learning rate), improved memory retention (e.g., lower weight decay and momentum), and larger memory storage (e.g., more hidden units). The greater learning rate for networks that reflected the children who passed the memory control enabled them to encode the first-order correlations within the allotted training trials. In contrast, the lower learning rate for networks that reflected the children who failed the memory control was insufficient to allow these networks to register the first-order correlations during training.

Combined with a higher learning rate, the lower weight decay and momentum and a larger set of hidden units allowed the networks with more robust information processing capacities to form an internal (hidden) representation that captured the first-order correlations (e.g., yellow sticker and red cube, red cube and machine activation) as well as the second-order correlation between them (e.g., yellow stickers and machine activation). Crucially, this ability to form a hidden representation that captured both the first- and second-order correlations served to facilitate second-order correlation learning only in those children who passed the memory control.

General discussion

The current investigation demonstrated that young children could encode second-order correlations to make causal inferences, but only if they recalled the feature–block pairings from the static phase. The computational model showed that an associative-learning mechanism combined with sufficient information-processing capacities is critical for registering second-order correlations among causal stimuli. Together, the findings from this study extend previous findings on this topic by showing that children with sufficient information-processing capacities can use second-order correlation learning to infer causal relations.

Children's developing information-processsing abilities may also explain why second-order correlation learning emerges at different ages in studies on this topic (e.g., Rakison & Benton, 2019; Yermolayeva & Rakison, 2016). For example, in Cuevas et al.'s (2006) study, 6-month-olds showed second-order correlation learning, but only after extended exposure to the stimuli. In contrast, children in the current study were given two brief exposures to the first-order correlations that constituted the second-order correlation. This difference in familiarization time might have affected the current children's ability to recall the initial association. Critically, age was not a critical factor for encoding the second-order correlations in the current study. Instead, the results indicated a relation between children's age and their developing memory demands. The computational model simulated how this difference in memory capacity might contribute to differences in second-order correlation learning. This suggests that second-order correlation is an important learning mechanism that depends on children's developing information-processing capacities. Importantly, the "age" of the model was not a relevant factor in the simulation; the results held regardless of whether the model was trained for 15 or 250 epochs.

More generally, these results suggest that children might be using second-order correlation learning to interpret correlational data in causal ways. To illustrate, consider a study by Saxe, Tzelnic, and Carey (2007), where 9-month-olds were habituated to alternating events in which a bean bag was tossed either from behind an opaque screen located on the left-hand side of a stage or from behind an opaque screen on the right-hand side of the stage. The screens were then lowered to reveal a hand or puppet with human-like eyes on one side of the stage and a toy truck on the other side of the stage. Saxe et al. (2007) found that 9-month-olds looked longer at the event in which the beanbag ostensibly was tossed from the side on which the toy truck was located. They interpreted these data as suggesting that infants understand that people, but not inanimate objects, are agents that can cause ballistic motion. An alternative view is that infants' conception of people as agents may well have derived from encoding two separate correlations, namely that people have hands and eyes and people are self-propelled. Based on these separate associations, infants might have understood that hands/eyes and ballistic motion are correlated through a second-order association. Future studies should look carefully at the explanatory nature of second-order correlations as a mechanism for children's causal inference, particularly in interpreting investigations with infants.

To conclude, the current experiment demonstrates that second-order correlation learning for causal stimuli is present by 2 years of age and that encoding the first-order relations—on which the second-order correlations are built—is critical for second-order correlation learning. The current results also suggest that second-order correlation learning may be grounded in associative processes and represents a powerful mechanism by which infants and children can form broad generalizations that span domains and content areas.

Acknowledgments

This work was supported by a National Science Foundation (NSF) grant (1661068) to D.M.S. We thank the children and their caregivers for their participation.

Data availability

The data that support the findings of this study are available from the corresponding author (dbenton2@swarthmore.edu) upon reasonable request.

References

Cuevas, K., Rovee-Collier, C., & Learmonth, A. E. (2006). Infants form associations between memory representations of stimuli that are absent. *Psychological Science, 17*, 543–549.
Cohen, L. B., & Amsel, G. (1998). Precursors to infants' perception of the causality of a simple event. *Infant Behavior and Development, 21*, 713–731.
Gopnik, A., & Sobel, D. M. (2000). Detecting blickets: How young children use information about novel causal powers in categorization and induction. *Child Development, 71*, 1205–1222.
Haith, M. M. (1993). Future–oriented processes in infancy: The case of visual expectations. In C. Granrud (Ed.), *Carnegie Mellon Symposia on Cognition: Visual perception and cognition in infancy* (pp. 235–264). Hillsdale, NJ: Lawrence Erlbaum.
Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition, 83*, B35–B42.
Lee, M. D., & Wagenmakers, E. J. (2014). *Bayesian data analysis for cognitive science: A practical course*. New York: Cambridge University Press.
Leslie, A. M., & Keeble, S. (1987). Do six-month-old infants perceive causality?. *Cognition, 25*, 265–288.
Luchkina, E., Sommerville, J. A., & Sobel, D. M. (2018). More than just making it go: Toddlers effectively integrate causal efficacy and intentionality in selecting an appropriate causal intervention. *Cognitive Development, 45*, 48–56.
Meltzoff, A. N., Waismeyer, A., & Gopnik, A. (2012). Learning about causes from people: Observational causal learning in 24-month-old infants. *Developmental Psychology, 48*, 1215–1228.
Oakes, L. M., & Cohen, L. B. (1990). Infant perception of a causal event. *Cognitive Development, 5*, 193–207.
Rakison, D. H., & Benton, D. T. (2019). Second-order correlation learning of dynamic stimuli: Evidence from infants and computational modeling. *Infancy, 24*, 57–78.
Rakison, D. H., & Krogh, L. (2012). Does causal action facilitate causal perception in infants younger than 6 months of age?. *Developmental Science, 15*, 43–53.
Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274*, 1926–1928.
Saxe, R., Tzelnic, T., & Carey, S. (2007). Knowing who dunnit: Infants identify the causal agent in an unseen causal interaction. *Developmental Psychology, 43*, 149–158.
Sobel, D. M., & Kirkham, N. Z. (2006). Blickets and babies: The development of causal reasoning in toddlers and infants. *Developmental Psychology, 42*, 1103–1115.
Waismeyer, A., & Meltzoff, A. N. (2017). Learning to make things happen: Infants' observational learning of social and physical causal events. *Journal of Experimental Child Psychology, 162*, 58–71.
Waismeyer, A., Meltzoff, A. N., & Gopnik, A. (2015). Causal learning from probabilistic events in 24-month-olds: An action measure. *Developmental Science, 18*, 175–182.
Walker, C. M., & Gopnik, A. (2014). Toddlers infer higher-order relational principles in causal learning. *Psychological Science, 25*, 161–169.
Walker, C. M., Lombrozo, T., Williams, J. J., Rafferty, A. N., & Gopnik, A. (2017). Explaining constrains causal learning in childhood. *Child Development, 88*, 229–246.
Woodward, A. L. (1998). Infants selectively encode the goal object of an actor's reach. *Cognition, 69*, 1–34.
Yermolayeva, Y., & Rakison, D. H. (2016). Seeing the unseen: Second-order correlation learning in 7- to 11-month-olds. *Cognition, 152*, 87–100.