Can 1- and 2-year-old toddlers learn causal action sequences?

Anonymous CogSci submission

Abstract

Toddlers can learn cause-effect relationships between a single action and its outcome. However, causality is often more complex. We investigate whether toddlers (12- to 35-month-olds) can learn that a sequence of two actions is causally necessary, from observing the actions of an adult demonstrator. In Experiment 1, toddlers saw evidence that performing a two-action sequence (AB) on a puzzle-box was necessary to produce a sticker, and evidence that B alone was not sufficient. Toddlers were then given the opportunity to interact with the box and retrieve up to 5 stickers. Toddlers had difficulty learning that a two-action sequence is causally necessary, with the ability to do so improving with age. In Experiment 2, toddlers saw evidence that performing a single action (B) was sufficient to produce an effect (and a sequence is not causally necessary). Toddlers performed fewer sequences in Experiment 2, suggesting some sensitivity to the sequential causal structure.

Keywords: Causal reasoning, Cognitive development, Social learning, Sequence learning

Introduction

The physical and social worlds are governed by a variety of simple and complex causal relationships. Sequences of multiple actions that need to be performed in a specific order to achieve a goal are common in our everyday routines. Take for instance getting a coke from a vending machine. What are the steps you would take to retrieve this item? Do you find the price and code for the item on the machine first? Do you enter the code before you insert your coins into the slot? What if the item gets stuck in the machine? Understanding that a sequence of actions is necessary to bring about an effect, and that these steps must be performed in a particular order (e.g., you must enter the item code before inserting the coins) enables us to predict subsequent events and to intervene on and manipulate our environment to achieve our goals. It has been suggested that our ability to encode sequential information may set us apart from other species (Ghirlanda et al., 2017).

From a young age, children use causality as a guiding principle for learning about the mechanisms of their environment, their own behaviour and that of others (see Muentener & Bonawitz, 2017; Sobel & Legare, 2014 for recent reviews). Research has shown that from preschool age, children are able to understand many of the principles governing causal relationships, such as covariation—that causes and effects co-occur, with causes predicting effects (Shultz & Mendelson, 1975; Mendelson & Shultz, 1976; Irwin, 1996), and temporal priority—that causes must precede their events in time (Bullock & Gelman; 1979; Rankin & McCormack, 2013).

Even 12- to 24-month-olds show relatively complex understanding of cause-effect scenarios, including understanding conditional independence (Sobel and Kirkham, 2006); applying causal rules based on abstract relations (Walker & Gopnik, 2014) and higher-order generalisations (Sim & Xu, 2017); and the ability to apply learned causal functions to solve novel problems (Goddu & Gopnik, 2020). In some causal reasoning tasks, toddlers even outperform preschoolers (e.g., Walker & Gopnik, 2014).

Young children, including toddlers, can learn causal structure by observing and copying more knowledgeable individuals (e.g., Meltzoff et al., 2012). Studies of deferred imitation suggest 11.5- and 13.5-month-old infants can remember order information for short novel sequences (e.g., Bauer & Mandler, 1992), but recall is better when sequences involve enabling causal relations—i.e., when one action enables the next to be performed. In enabling situations, the actions can only be performed in one specific order (e.g., unlocking a box enables the lid to be lifted; Bauer, 1992).

Further evidence that enabling sequences are easier for young children to reproduce comes from a study by Brugger et al. (2007), who demonstrated two actions leading to an interesting effect to 14- to 16-month-olds. When a 2-action sequence was causally necessary due to an enabling causal relation, toddlers were more likely to copy the sequence, than when the first action was not causally necessary. However, even in the enabling scenario, only 29% of participants reproduced the demonstrated sequence, whereas 39% performed either the first or the second action, but not both.

Similarly, Carpenter et al. (1998) found that 14- and 18-month-old infants were able to readily reproduce a single demonstrated action out of two possible actions, but when presented with a two-action sequence, only 6/20 infants spontaneously reproduced the demonstration in the correct order. These findings suggest that infants and toddlers can reproduce short action sequences—especially if they involve enabling causal relations—but they do not do so reliably.

The extent to which even older children can reliably copy a sequence of actions in the correct order is not entirely clear. Studies of overimitation with preschoolers have shown that they can copy multiple novel actions to achieve a goal (e.g., Horner & Whiten, 2005; Lyons et al., 2007). However, whether these actions are performed in the correct sequence is not often explicitly coded for, and there is evidence that 4- and 5-year-olds can struggle to acquire temporal information for arbitrary action sequences (Loucks & Price, 2019).

Additional work has shown that preschoolers readily reproduce a 2-action sequence in the correct order, when it is ambiguous whether the first action is causally necessary to enable the second one (Anonymized, under revision). Specifically, when 3- to 5- year olds watched a demonstrator perform a sequence of two actions (AB) on a causally opaque puzzle-box, that led to a desirable effect (E, a sticker popping out of the box), they faithfully copied the sequence, performing actions in the correct order. In contrast, when 18- to 30-month-old toddlers watched the same demonstration they did not reproduce the sequence; instead, they tended only to perform the second action (B) from the sequence (Anonymized, 2020a). This pattern of faithful copying—or overimitation—increasing across early childhood fits with evidence from other studies (Hoehl et al., 2019).

Additional research using a similar paradigm suggests that a recency effect does not account for why toddlers in Anonymized (2020a) only reproduced the second action that they saw. When 12- to 35-month-olds watched a demonstrator perform A, following which a sticker dispensed (effect E), following which a second action (B) was performed, they were significantly more likely to (correctly) manipulate A than B (Anonymized, under review)—the opposite of what would be predicted by a recency effect.

These previous findings raise the possibility that toddlers have difficulty understanding that sequences of actions can be causally necessary—potentially because they have a strong prior expectation that single outcomes have a single cause (Anonymized, 2020a). However, in the study where toddlers saw AB-E demonstrated, and primarily intervened on just B, the demonstration was purposefully ambiguous in terms of whether the AB sequence was necessary to produce the effect (and in fact, only action B was necessary). In the present study, we asked whether toddlers could learn the correct sequential causal structure in a simple puzzle-box paradigm, when the demonstrations they see provide unambiguous evidence regarding whether a sequence is necessary or unnecessary.

In Experiment 1, we asked whether toddlers grasp that a sequence of two actions is causally necessary when they see unambiguous evidence that a single action is not sufficient to produce the desirable outcome, but that the sequence does produce the outcome. 12- to 35-month-olds observed an experimenter manipulate a puzzle-box by performing action A (e.g., pushing a button) followed by action B (e.g., sliding a handle) to produce an effect E (dispensing a sticker). They also observed the demonstrator perform only action B, which did not lead to a sticker being dispensed. Toddlers then had the chance to interact with the box themselves and retrieve up to five stickers. If toddlers can infer the correct causal structure from the observed evidence (i.e., that the AB sequence is necessary), then they should act on A first, followed by B. If toddlers perform A followed by B in quick succession (like the demonstrator), then this would be particularly compelling evidence that they grasp that the sequence is causally necessary.

In comparison, in Experiment 2, 12- to 35-month-olds saw evidence that only the single action B was causally necessary. Specifically, they saw a demonstration that both the 2-action sequence (AB) and the single action B led to a sticker being dispensed. If toddlers can learn causal sequences then they should produce AB sequences in Experiment 1 and should produce more in Experiment 1 than Experiment 2, even though they have seen AB and B performed equally in both experiments. On the other hand, if they have difficulty learning sequences and inferring that a sequence is necessary, then they should perform relatively few sequences and have more difficulty successfully retrieving stickers in Experiment 1 than Experiment 2. Finally, if they just copy the single final action that they saw precede the effect, then they should act on B first equally in both cases.

Experiment 1

Participants

Fifty-five 12- to 35-month-old children were included in Experiment 1 ($M_{age,\ months}$ = 23.89). Nineteen additional participants were tested but excluded due to: equipment failure (N = 8), experimenter error (N = 2), parental or sibling interference (N = 6), and did not complete session due to distraction (N = 3). An additional 12 participants completed the experiment but did not interact with the puzzle-box and therefore did not provide any data.

Methods and Materials

Stimuli

A customized wooden puzzle-box was used as in Anonymized (2020a). The box was equipped with three different-coloured, interchangeable fronts, each with two distinct actions at opposite sides and a reward dispenser at the centre below the actions and equidistant from them. The assignment of front panels and actions as A and B were counterbalanced across participants. A rotating motor with seven wells was placed inside of the puzzle-box and covertly activated with a remote by the experimenter to dispense a sticker contained in an Eppendorf tube into a tray.

The puzzle-box was placed on a low table to ensure accessibility for the young children. Cameras recorded from two angles to capture the child's observation during the demonstration phase and manipulation of the box during the action phase. During the sessions, caregivers were either present or seated near the child, at off-site testing, or in an observation room, watching the participant through a one-way mirror, at in-lab testing.

Procedure

Acclimatization. An adult female experimenter (E1) interacted with the child, while turned away from the puzzle-box, to acclimate the child to the testing environment.

Demonstration. A second adult female experimenter (E2), in the testing area, did not interact with or acknowledge the child, to minimize the impact of social cues on actions. Once E1 and the child turned towards E2, E1 said, "Oh, it looks like E2 is using the room right now, let's wait for our turn over here, we can watch!" Without acknowledging E1 or the participant, E2 performed two demonstrations. In one demonstration, E2 manipulated A and then B, following which a sticker was dispensed from the puzzle-box (A-B-E). E2 picked up the sticker and said, "Oh, a sticker!" before placing it back into the tray. In the other demonstration, E2 manipulated B and then no sticker was dispensed (B-No Effect). E2 placed a hand in the empty tray and said, "Oh, no sticker." E2 then repeated these same two demonstrations in the same order, which was counterbalanced across participants (half saw A-B-E first, and half saw B-No Effect first). If a toddler missed demonstration, as signalled by E1, it was repeated until they had seen each type twice. Following the demonstrations, E2 acted busy again before suddenly noticing the child saying, "Oh, I'm all done here, you can have a turn!" and leaving the testing area.

Child Action Phase. E1 and the child approached the puzzle-box. The participant was able to interact with it to receive up to five stickers. A sticker was dispensed when the participant manipulated A and then B, irrespective of the time elapsed between these actions. If the participant did not spontaneously interact with the puzzle-box, E1 provided neutral encouragement such as, "It's your turn, you can try anything!" Once five stickers were dispensed, E2 returned to the testing area and said, "You got all the stickers!", and the session ended. If the participant did not interact with the box for >2 min, the session was also ended.

Data Scoring and Analysis

All sessions were coded live and then re-coded from footage. Each child could activate the puzzle box up to 5 times. An activation ended either when the sticker was released or, if the child failed to successfully retrieve a sticker on that activation, when the session ended (the latter could only occur on the final activation performed by a given child). Each child was given a score of 0-5 for the number of successful activations. Within each activation, each action was noted in the order in which it was manipulated by the child (e.g. AAAAB). A single action was defined as a distinct touch (e.g. pushing the button in and then letting go) or a continuous hold of the action with a distinct motion (e.g. holding the button and pressing it in and out without pausing). From this action stream we coded whether toddlers touched *A or B first* for each activation of the puzzle-box. We also coded *strict A-B sequences,* defined as A and then B being manipulated within 5 seconds (with no preceding touches to B and no intervening touches to A), to give a fairly conservative measure of toddlers sequence reproduction (though still reasonably generous as the demonstration of the AB sequence was more rapid than this). Finally, since toddlers might perform a sequence less fluently than older children and adults, we also coded *loose A-B sequences*, defined as a first touch to A followed by B within any amount of time (and allowing for more than one touch of A). Since not all toddlers performed 5 activations, we analyzed all the DVs except number of successful activations as proportions, however analyzing the raw number of each sequence type does not significantly change results.

Results and Discussion

In Experiment 1, toddlers successfully activated the puzzle box on average 3.35 out of 5 possible times, however the number of times toddlers successfully activated the box increased with age ($\beta = 2.00$, $p < .001$, Figure 1a). When split by median age, one year olds ($M = 2.37$, $SD = 2.13$) were significantly less successful than 2 year olds ($M = 4.29$, $SD = 1.46$), $t(53) = 3.90$, $p < 0.001$. Toddlers also produced significantly more *strict AB sequences* (a sequence of A-B within 5 seconds) with increasing age ($\beta = .23$, $p < .001$, Figure 1b), and a median split showed that 2-year-olds produced significantly more strict sequences ($M = .27$, $SD = .33$) than 1-year-olds ($M=0.04$, $SD = .09$), $t(53) = 3.440.3$, $p <.001$. The production of *loose AB sequences* showed a similar but weaker trend, ($\beta = .16$, $p = .068$, Figure 1c), with 2-year-olds again producing more loose sequences ($M = .48$, $SD = .38$) than 1-year-olds ($M=0.29$, $SD = .35$), $t(53) = 1.93$, $p = 0.059$. We also examined toddler's *first touch* on each activation, and found that overall they were equally likely to first act on either A or B ($M=0.49$, $SD = .40$), $t(54) = 0.25$, $p = .81$, and that this was consistent across age groups ($\beta = .005$, $p = .0.95$).

Overall, when a sequence was causally necessary to produce an effect, children were increasingly likely to produce one with age. Two-year-old toddlers were more likely to be successful and use either a strict or loose AB sequence (strict: 27%, loose: 48%), compared to their 1-year-old counterparts (strict: 4%, loose: 29%). To further investigate the understanding of sequential causal structure in toddlers, In Experiment 2 we examined their behavior when they are given evidence that only a single action is causally necessary for the effect, even though a sequence of both actions has still been demonstrated.

Experiment 2

In Experiment 2, 1- and 2-year-olds saw evidence that only the single action B was causally necessary. Specifically, demonstrations of the 2-action sequence (AB) *and* the single action B led to a sticker being dispensed. If toddlers distinguish this evidence from that presented in Experiment 1 and can infer that in this case only B is necessary, then they should produce fewer AB sequences here, even though they have seen AB and B performed equally in both experiments. Given that younger toddlers in Experiment 1 found it challenging to activate the box by performing a sequence, they should be more successful here where a sequence is not necessary. Further, if toddlers are sensitive to the necessity of the sequence in Experiment 1 and its lack of necessity in Experiment 2, then they should produce fewer sequences in Experiment 2.

Participants

Fifty-four new 12- to 35-month-old children participated in Experiment 2 ($M_{age,\ months}$ = 24.36). Seventeen additional participants were tested but excluded due to: equipment failure (N = 4), experimenter error (N = 3), parental or sibling interference (N = 9), and did not complete session due to distraction (N = 1). An additional 3 children watched the demonstrations but did not interact with the puzzle-box.

Methods and Materials

All Methods and Materials, including Stimuli, Procedure, and Data Collection and Analysis, were the same as Experiment 1, except for the following differences in the Procedure:

Demonstration. As in Experiment 1, E2 manipulated A and then B, following which a sticker was dispensed (AB-E). Next, E2 manipulated B only, and another sticker was dispensed (B-E). The same two demonstration were repeated in the same order, and counterbalanced across participants, where half saw AB-E first and the other half saw B-E first. Thus, toddlers saw evidence that only the single action B was causally necessary.

Child Action Phase. In contrast to Experiment 1, the puzzle box dispensed a sticker when the participant acted on B.

Results and Discussion

In Experiment 2, age did not significantly influence toddlers' success at activating the puzzle-box (β = .56, p = .08). Toddlers successfully activated the puzzle box on average 4.4 out of 5 possible times. A median split showed that 2-year-olds (M = 4.75, SD = .52) tended to activate the box more times than 1-year-olds (M = 4.08, SD = 1.72; Figure 1a), but the difference was not significant (t(52) = 1.98, p = .053). As in Experiment 1, older toddlers produced significantly more *strict AB sequences* than younger toddlers, β = .14, p = .03, (Figure 1b), but unlike in Experiment 1 a median split revealed no significant difference between 1-year-olds (M = .12, SD = .23) and 2-year-olds (M = .22, SD = .30), t(52) = 1.35, p = .18. The production of *loose AB sequences* did not significantly change with age, β = .11, p = .12 (Figure 1c). We also examined toddler's *first touch* on each activation, and found that, unlike in Experiment 1, toddlers were significantly more likely to first act on action B, M = .74, SD = 0.32, (t(53) = 5.43, p < .001, and that this was consistent across age groups, β = .001, p = .98.

In Experiment 2, toddlers were successful at activating the puzzle-box. Toddlers activated the puzzle-box most frequently through the usage of the B action on its own (74%), compared to using a loose AB sequence (22%), with other sequences comprising the remaining 4%. Nonetheless, as in Experiment 1, the tendency to produce strict AB sequences increased with age, even though here a sequence was not causally necessary.

Comparison of Experiments 1 and 2

To examine the extent to which toddlers may have differentiated the evidence observed in Experiments 1 and 2, we compared their performance between experiments for our main DVs. For the *number of successful puzzle-box activations*, there was a significant main effect of experiment F(1, 105) = 12.88, p < .001 as well as age, F(1, 105) = 25.08, p < .001, and a significant interaction between these factors, F(1, 105) = 7.99, p < .001. Older toddlers successfully activated the puzzle-box significantly more times than younger toddlers in Experiment 1, whereas success was high across the age range in Experiment 2 (Figure 1a).

For *strict AB sequences*, there was a significant main effect of age, F(1, 105) = 10.92, p < .001, such that older toddlers produced significantly more strict AB sequences than younger toddlers (Figure 1b). There was no main effect of experiment (F(1,105) = .19, p = .66) and no interaction (F(1,105) = 1.54, p = .22). For *loose AB sequences* there was a significant main effect of experiment, F(1, 105) = 7.73, p = .006, as well as age, F(1, 105) = 6.00, p = .016, but no interaction, F(1,105) = .17, p = .68. Toddlers were more likely to produce loose sequences in Experiment 1 than Experiment 2, and older toddlers were significantly more likely than younger toddlers to produce loose AB sequences in both experiments (Figure 1c). Finally, for the first touch on each activation, there was a main effect of experiment, F(1, 105) =10.19, p = .002, but no effect of age F(1, 105) = .003 p = .96, and no interaction (F(1,105) = 0.001, p = .98).

Taken together, this comparison of performance between Experiments 1 and 2 shows that toddlers—especially 1-year-olds—found it easier to activate a puzzle-box where only a single action (B) was required (Experiment 2), compared with a puzzle box where a 2-action sequence (AB) was causally necessary (Experiment 1). The extent to which toddlers produced strict AB sequences increased with age and did not differ between experiments, suggesting that, according to this measure, they did not differentiate the causal structures in Experiment 1 vs. Experiment 2. Nonetheless, our loose AB sequence measure did reveal a difference in behavior between experiments: performing A followed by B was more common in Experiment 1 where the sequence was causally necessary. In addition, toddlers were more likely to act first on action A in Experiment 1, where A was necessary, than in Experiment 2 where it was not. As with the strict sequence measure, performance of loose sequences increased with age, providing evidence that toddlers, and particularly 1-year-olds, may find it challenging to grasp and/or perform causal action sequences.

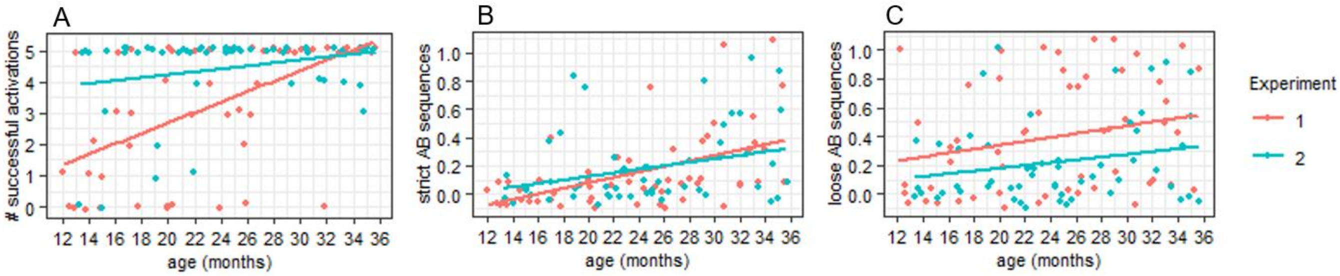

Figure 1: Linear regression of: (a) mean successful activations, (b) proportion of activations preceded by a strict AB sequence, (c) proportion of activations preceded by a loose AB sequence, as a function of age in Experiment 1 and 2.

General Discussion

We investigated whether 1- and 2-year-olds can learn through observation that a sequence of two actions (AB) is causally necessary, when they see unambiguous evidence that just a single action (B) is insufficient to produce a desirable effect (Experiment 1). We also investigated whether toddlers of the same age could learn that a sequence was *not* necessary when they saw evidence that both a 2-action sequence (AB) and single action (B) were equally effective (Experiment 2). We compared performance in these two experiments to see whether toddlers behaved differently depending on the evidence they saw, even though the experimenter produced the same sets of actions in both cases.

Our results suggest that the ability to learn causal sequences may develop over early childhood. When a sequence was causally necessary in Experiment 1, 2-year-olds were more successful at activating the puzzle-box and produced more strict AB sequences than 1-year-olds. However, the production of strict sequences (a first touch to A followed by B within 5 seconds, comparable to what was demonstrated by the experimenter) were relatively rare across the age range, suggesting that learning a short causal sequence remains challenging in the third year of life. Loose AB sequences (a first touch to A followed by B within any amount of time and allowing for more than one touch of A) were more common, but still produced more by older than younger toddlers.

In Experiment 2, where action B alone was causally effective, toddlers across the age range were more successful at activating the puzzle-box than their counterparts in Experiment 1. Loose AB sequences were more common in Experiment 1, and toddlers in Experiment 1 were significantly more likely to act on A first, suggesting that toddlers could differentiate causal relevance based on evidence provided in the demonstration to some extent. This indicates that sequential causal learning improves with age and that toddlers find it easier to learn causal structure when a single action is causal, rather than a 2-action sequence.

In a previous study (Anonymized, 2020a), when toddlers were shown an AB sequence that resulted in a desirable outcome where the necessity of action A was ambiguous, they tended to omit A and only perform B. The current study builds on this work by showing that when they see unambiguous evidence that a sequence is necessary (because B alone is ineffective) toddlers still rarely copy an AB sequence.

Toddlers may prefer single-action causes due to a strong prior belief that 2-action causes are unlikely, as suggested by previous work (e.g., Carpenter et al., 1998, Brugger et al., 2007; Anonymized, 2020). In Experiment 1, the toddlers may have performed AB sequences infrequently, despite evidence of causal relevance, as they *a priori* believe multi-action sequences to be unlikely causes.

Given evidence that even preschoolers can struggle to recall temporal information for action sequences (Loucks & Price, 2019) it is also possible that toddlers found it difficult to remember the order of the two actions they saw demonstrated, particularly as there were multiple demonstrations of different types in each experiment, which they may have been challenging to parse.

The behavior of toddlers may have been influenced by the features of the experiment, in addition to/instead of their (in)ability to grasp that a 2-action sequence can be causally necessary. For example, though the experimenter in the study acted intentionally, they did not present as knowledgeable about the puzzle-box, to avoid the use of a pedagogical context and inadvertently encourage overimitation, or potentially override statistical information with ostensive cues (e.g., Marno & Csibra, 2015). Therefore, toddlers may have viewed the experimenter as exploring the box and wondering why it failed, rather than explicitly teaching them how to operate it. In addition, the demonstration of the single action B in Experiment 1 may have suggested to toddlers that this action was expected by the demonstrator to be causally effective, but it failed, which may have prompted toddlers to explore the puzzle-box, rather than to act efficiently. In the future, it would be interesting to see how the incorporation of explicit teaching cues might influence toddlers' causal inferences in this task.

In Experiment 1, where an AB sequence was causally necessary, we anecdotally observed that a number of toddlers performed the actions BAB consecutively, which is also a valid causal sequence (though the first B is unnecessary). Further analysis of the timing of and duration between toddlers' actions will be conducted to disentangle alternative potential explanations for this unexpected behavior. For example, it is possible that some toddlers believed that BAB was the causally necessary sequence (in which case we would expect short durations between actions). Alternatively, this pattern could emerge if toddlers believed that B alone was causal, tried it and found it to be ineffective, so tried A, and then at some point performed B again, which activated the box (due to a loose AB sequence being executed). This could then reinforce the belief that B was the sole cause. In this case we would expect quite different action execution timings.

In the present study, we were specifically interested in causal sequence learning in the context of a causally opaque puzzle-box. However, given that prior evidence suggests that sequences involving enabling causal relations are easier for toddlers to master (Bauer, 1992; Brugger et al., 2007), it would be interesting to investigate whether the ability to learn causal sequences improves when it is visibly obvious that a sequence is causally necessary (e.g., pulling the lever on the box moves the sticker into a position that enables the dial to push it down into the dispenser).

Given that even older 2-year-olds were not at ceiling in the current study, gathering further information about the developmental trajectory of the ability to learn causal sequences seems warranted. To this end, ongoing work is testing preschoolers with the existing task.

In conclusion, although toddlers show relatively sophisticated causal reasoning skills in some tasks, including learning through observing others (e.g., Meltzoff et al., 2012; Walker & Gopnik, 2014), this study suggests that, at least in the context of a causally opaque puzzle-box, the ability to learn sequential causal structure via observation may develop across early childhood. Young toddlers may struggle to grasp that a sequence of actions can be causal, or at least find causal sequences highly implausible.

References

Bauer, P. J. (1992). Holding it all together: How enabling relations facilitate young children's event recall. *Cognitive Development*, *7*(1).

Bauer, P. J., & Mandler, J. M. (1992). Putting the Horse before the Cart: The Use of Temporal Order in Recall of Events by One-Year-Old Children. *Developmental Psychology*, *28*(3).

Brugger, A., Lariviere, L. A., Mumme, D. L., & Bushnell, E. W. (2007). Doing the Right Thing: Infants' Selection of Actions to Imitate From Observed Event Sequences. *Child Development*, *78*(3).

Bullock, M., & Gelman, R. (1979). Preschool children's assumptions about cause and effect: Temporal ordering. *Child Development*, *50*(1).

Anonymous (under review a).

Carpenter, M., Akhtar, N., & Tomasello, M. (1998). Fourteen- through 18-month-old infants differentially imitate intentional and accidental actions. *Infant Behavior and Development*, *21*(2).

Ghirlanda, S., Lind, J., & Enquist, M. (2017). Memory for stimulus sequences: a divide between humans and other animals? *Royal Society Open Science*, *4*(6), 161011.

Goddu, M. K., & Gopnik, A. (2020). Learning what to change: Young children use "difference-making" to identify causally relevant variables. *Developmental Psychology*, *56*(2).

Hoehl, S., Keupp, S., Schleihauf, H., McGuigan, N., Buttelmann, D., & Whiten, A. (2019). 'Over-imitation': A review and appraisal of a decade of research. *Developmental Review*, *51*.

Horner, V., & Whiten, A. (2005). Causal Knowledge and Imitation/Emulation Switching in Chimpanzees (Pan troglodytes) and Children (Homo sapiens). *Animal Cognition*, *8*.

Irwin, K. C. (1996). Children's Understanding of the Principles of Covariation and Compensation in Part-Whole Relationships. *Journal for Research in Mathematics Education*, *27*(1).

Loucks, J., & Price, H. L. (2019). Memory for temporal order in action is slow developing, sensitive to deviant input, and supported by foundational cognitive processes. *Developmental Psychology*, *55*(2).

Lyons, D. E., Young, A. G., & Keil, F. C. (2007). The hidden structure of overimitation. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(50).

Marno, H., & Csibra, G. (2015). Toddlers favor communicatively presented information over statistical reliability in learning about artifacts. *PloS One*, *10*(3), e0122129.

Meltzoff, A. N., Waismeyer, A., & Gopnik, A. (2012). Learning about causes from people: Observational causal learning in 24-month-old infants. *Developmental psychology*, *48*(5), 1215.

Mendelson, R., & Shultz, T. R. (1976). Covariation and temporal contiguity as principles of causal inference in young children. *Journal of Experimental Child Psychology*, *22*(3).

Muentener, P., & Bonawitz, E. (2017). *The development of causal reasoning*. In M. R. Waldman (Ed.), The Oxford Handbook of Causal Reasoning. New York, NY: Oxford University Press.

Rankin, M. L., & McCormack, T. (2013). The temporal priority principle: At what age does this develop? *Frontiers in Psychology*, *4*.

Shultz, T. R., & Mendelson, R. (1975). The use of covariation as a principle of causal analysis. *Child Development*, *46*(2).

Sim, Z. L., & Xu, F. (2017). Infants preferentially approach and explore the unexpected. *The British Journal of Developmental Psychology*, *35*(4).

Sobel, D. M., & Kirkham, N. Z. (2006). Blickets and babies: The development of causal reasoning in toddlers and infants. *Developmental Psychology*, *42*(6).

Sobel, D. M., & Legare, C. H. (2014). Causal learning in children. *WIREs Cognitive Science*, *5*(4).

Anonymous (2020a).

Anonymous (2020b)

Anonymous (under review b)

Walker, C. M., & Gopnik, A. (2014). Toddlers Infer Higher-Order Relational Principles in Causal Learning. *Psychological Science*, *25*(1).