Toddlers and Adults Simultaneously Track Multiple Hypotheses in a Causal Learning Task

Mariel K. Goddu (goddu@berkeley.edu)
Department of Psychology, Tolman Hall
Berkeley, CA 94618 USA

Caren M. Walker (carenwalker@ucsd.edu)
Department of Psychology, McGill Hall #0109
9500 Gilman Drive, La Jolla, CA 92093 USA

Abstract

Research on the development of future hypothetical and counterfactual thinking suggests that children as old as five may be unable to consider multiple, equally probable possibilities simultaneously. Yet, a large literature on the development of causal reasoning suggests that much younger children are able to generate, evaluate, and test causal hypotheses, often by integrating information about several candidate causes at once. The current research seeks to bridge these two bodies of research. In three experiments, adults and toddlers (18–30 months) observe a sequence of evidence that is equally consistent with two hypotheses, each occupying a different level of abstraction (individual vs. relational). Results suggest that learners generate more than one potential cause, hold both in mind, and flexibly apply the appropriate hypothesis to inform their inferences at test. Findings challenge previous suggestions that much older children fail to consider multiple, equally probable possibilities.

Keywords: cognitive development; causal reasoning; counterfactual thinking; epistemic uncertainty

Introduction

As both children and adults encounter new data, they update their hypotheses, revise their beliefs, and adjust their behavior accordingly. This view—perhaps the central tenet of modern theories of cognitive development—assumes that learners have a "hypothesis space" populated by many possibilities with varying likelihoods (Gopnik et al., 2004). From this perspective, learning itself may be interpreted as a process of assessing the probabilities of multiple possibilities.

In many causal learning studies, participants must use patterns of statistical contingency to determine which causes or causal rules best account for observed effects. Strikingly, even toddlers are often successful in these tasks (e.g., Gopnik, Sobel, Schulz, & Glymour, 2001; Gopnik & Sobel, 2000; Gweon & Schulz, 2011; Meltzoff, Waismeyer, & Gopnik, 2012). Yet puzzlingly, studies that ask much older children to make predictions about scenarios with multiple possible outcomes often find that they are unable to consider those possibilities simultaneously. Some researchers have posited that this is indicative of a general deficit in children's ability to reason about multiple possibilities at once (e.g., Beck, Robinson, Carroll, & Apperly, 2006; Redshaw & Suddendorf, 2016). However, causal learning tasks, which require participants to track and use causal rules rather than verbally predict or prepare for possible outcomes, may shed new light on this claim.

When two hypotheses accord equally well with the evidence observed in a causal task, do learners consider both possibilities simultaneously? Or, do they attend to only one? Here we investigate whether adults and very young children—toddlers aged 18 to 30 months—consider multiple hypotheses simultaneously in a simple causal reasoning task. Instead of assessing children's ability to verbally report or behaviorally prepare for multiple possible outcomes, the current task taps an early-emerging ability to generate and evaluate multiple possibilities as causal hypotheses. This work therefore represents a first step toward integrating two related, but hitherto distinct (and often conflicting) lines of research. Can causal learning illuminate the development of modal reasoning?

Reasoning about multiple outcomes

Previous research investigating the development of counterfactual thinking and reasoning about the outcome of future hypotheticals suggests that the ability to simultaneously track multiple hypotheses may not emerge until the early school years. This research has strongly suggested that preschoolers and even early school-aged children have difficulty considering multiple, equally probable possibilities simultaneously. Instead, these studies find that children tend to fixate on a single possibility (e.g., Beck, Robinson, Carroll, & Apperly, 2006; Ozturk & Papafragou, 2015; Rafetseder, Cristi-Vargas, & Perner, 2010; Redshaw & Suddendorf, 2016; Robinson, Rowley, Beck, Carroll, & Apperly, 2006).

Many of these tasks require children to learn rules about novel systems and make verbal reports about their predictions. For example, in several studies, children were required to infer that an object (e.g., a toy mouse) could emerge from *either* of two openings at the end of a forked tube, depending on which type of card was drawn at random from a deck. Though 3- to 5-year olds could competently answer questions such as, "What if next time [the mouse] goes the other way? Where will [it] be?", they were unable to prepare for the (equally likely) future possibilities that the mouse might exit either tube when the outcome was undetermined (i.e., had yet to occur). In these cases, children placed only one mat for catching the mouse, rather

than the required two. The researchers interpret these findings as evidence that preschool-aged children cannot effectively acknowledge and prepare for multiple, as yet undetermined possibilities (Beck et al., 2006). Notably, though, this paradigm requires that children process a large amount of complex verbal information and logical rules (i.e., drawing cards coded by color and pattern to determine the mouse's path in a sequence, varying from trial to trial).

In a simplified, non-verbal version of the same task, Redshaw and Suddendorf (2016) found that 2.5-year-old children perform similarly to apes in a task that requires catching a ball dropped down a forked tube. Instead of extending two hands to cover both openings, young children and apes extend only one hand. Although older children's performance improves over sequential trials, fewer than half of 3.5-year-olds make the appropriate response on the first trial. The authors conclude that these findings align with domain-general improvements in executive functioning and planning capabilities. This includes the ability to "metarepresent" that an imagined possibility (i.e., 'the ball will come out the *left* tube') may turn out to be incorrect.

Other types of tasks have assessed children's ability to reason about logical constraints and verbal rules that determine which of several outcomes may be possible in a given situation. For example, in an experiment conducted by Robinson and colleagues (2006), 5-year-olds were introduced to a house with three doors, out of which various blocks might emerge. The children were then shown two buckets of blocks: e.g., one composed entirely of black blocks, and the other composed of an equal amount of green and yellow blocks. Children were told that the experimenter would draw a block from each of the bins and push it through the doors in accordance with the following rules: if it is black, they would push it out the first door; if it is yellow, they would push it out the second door, and if it is green, they would push it out the third door. While children succeeded in placing a tray to catch the block underneath the first door when the experimenter drew from the (determinate) black block box, they failed to use multiple trays to prepare for the equally likely possibilities that a block could come out of *either* the second or third doors when the experimenter drew from the indeterminate green/yellow box. Like the children who placed only one mat for the mouse (Beck et al., 2006) or who extended only one hand to catch the ball (Redshaw & Suddendorf, 2016), these children also prepared for only one of two equally likely possibilities.

These and other instances in which children tend to treat an indeterminate state of affairs as determinate are known as "premature closure" (Ozturk & Papafragou, 2015; Robinson et al., 2006). When confronted with cases of epistemic uncertainty, children frequently fail to consider multiple, equally likely possibilities simultaneously. In the present experiments, we take an initial step at challenging this assumption.

Tracking Causal Possibilities

One reason to expect that even very young children may be capable of considering multiple hypotheses simultaneously is the fact that they are sophisticated causal reasoners. Causal reasoning skills emerge early in development across a wide range of tasks. Children as young as 16 to 24 months can observe patterns of statistical contingency between causes and effects and infer the causal properties of objects, intervene on causal systems to generate desired effects, and design novel causal interventions (e.g., Gopnik et al., 2001; Gopnik & Sobel, 2000; Gweon & Schulz, 2011; Meltzoff et al., 2012).

Recent research has demonstrated that toddlers' causal reasoning prowess extends to even more abstract forms of causality than previous studies have revealed. For example, previous research has found that toddlers are able to learn causal rules based on both *individual causation,* in which a single entity brings about an outcome, and on *relational causation,* in which the abstract relation that holds between two entities (e.g., sameness or difference) is responsible for an effect (Walker & Gopnik, 2014; Walker, Bridgers, & Gopnik, 2016; Walker & Gopnik, 2017). In these studies, toddlers are presented with a sequence of evidence that accords with a relational causal rule, such as "two different blocks placed on a machine will make it play music." This evidence includes two trials in which the machine is activated, interleaved with two trials in which pairs of "same" blocks fail to activate the machine. At test, toddlers are presented with a choice between two *novel* pairs of blocks, one same pair and one different pair (see Figure 1). Across studies, toddlers repeatedly succeed at inferring relational causes from this limited data (e.g., they readily select a novel pair of two "different" blocks to activate a machine to play music).

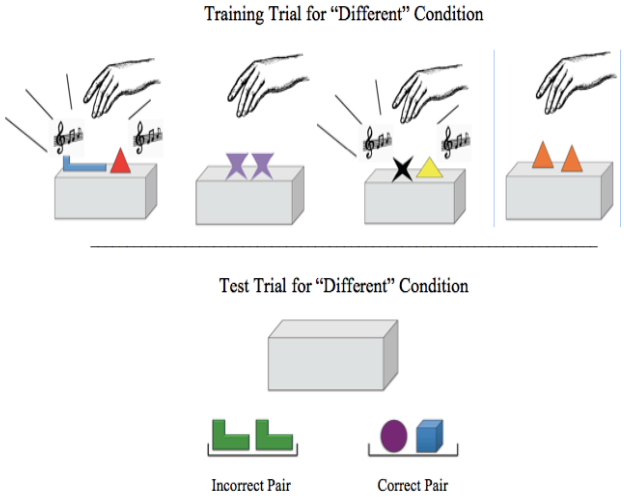

Figure 1. The causal relational match-to-sample task, "different" condition (Walker & Gopnik, 2014)

In both the classic causal learning studies that use individual causation and these newer studies that probe toddlers' ability to infer relational causation, participants are tasked with tracking several candidate causes and their corresponding effects (e.g., which blocks will or will not cause a machine to play music). As such, there is a sense in which these studies reflect young children's ability to consider multiple hypotheses at once. To our knowledge, however, there are as yet no studies in which toddlers are trained on ambiguous causal evidence that is *equally compatible with multiple hypotheses*.

Given that toddlers are sensitive to both individual and relational causation, this task presents a method for exploring our question: If participants are presented with evidence that supports both hypotheses equally, might they attend to both possible causes simultaneously?

Potential Challenges

While toddlers' success in learning a relational causal rule seems promising for their ability to detect and track *both* an individual level rule *and* a relational level rule simultaneously, there are a variety of reasons why the latter task may be more difficult. First, attending to individual objects and attending to the relations between them are processes often considered to trade off or to be in tension, particularly in young children. Indeed, in many relational reasoning tasks with preschoolers, attention to objects seems to preclude attention to relations (e.g., Christie & Gentner, 2010; Christie & Gentner, 2014). As such, it may be difficult for participants to hold both hypotheses in mind: doing so would require that they appreciate a potential individual cause (e.g., a block of a particular shape and color) both as an individual and as part of a pair, in relation to another. However, previous work indicates that toddlers' relational reasoning abilities may in fact be more readily available than those in preschoolers (Walker, Bridgers, & Gopnik, 2016). In particular, this study suggests that younger children's weaker priors on individual causation makes it easier for them to consider the relational option. This may mitigate concerns for learners at this age.

However, even if toddlers are immune to the type of object fixation or individual causation bias that older children typically display, toddlers' working memory capacity undergoes large developmental shifts between 12 and 36 months of age (e.g., Alp, 1994; Garon, Bryson, & Smith, 2008). It may therefore be difficult for children this young to hold multiple possible hypotheses "online." As a coping strategy, they may opt to fixate on only one of the two, or they may fail to generate a second hypothesis at all.

Despite these potential challenges, we set out to test whether toddlers consider multiple, equally sound hypotheses in light of ambiguous evidence. Given that the ability to track multiple hypotheses simultaneously is integral to a view of learning as a process in which learners constantly assess and re-assess possibilities, we predicted that this skill might be very early emerging.

The Current Experiments

In the current experiments, we presented participants with a revised version of the causal relational-match-to-sample procedure described above (Walker & Gopnik, 2014; Walker, Bridgers, & Gopnik, 2016; Walker & Gopnik, 2017). We capitalize on the fact that toddlers are able to learn both *individual causation* and *relational causation* hypotheses from evidence in this context. By presenting participants with an ambiguous sequence of evidence with which both individual and relational causation are equally compatible (i.e., one individual block appears twice, as a part of two "different" pairs—see Figure 2 and 3 below), we then compare the results of participants in two separate conditions who receive a forced choice between either individual or relational options to cause the machine to play music. We also test adults, as a basis for comparison.

In *Experiment 1*, we show that adults are equally likely to assume an individual causation hypothesis (e.g., "the black block makes the machine play music") as they are to assume a relational causation hypothesis (e.g., "two different blocks make the machine play music") from the ambiguous evidence.

In *Experiment 2*, we again presented adult participants with the ambiguous evidence, and then asked them to produce the effect at test. In contrast to *Experiment 1,* however, half of participants received a forced choice between the block indicated by the individual causation hypothesis and a decoy block (i.e., another single block that was associated with the effect). The other half received a forced choice between two novel pairs, one "same" and one "different." Adults in both conditions chose the correct possibility at above chance rates, suggesting that they were considering multiple possibilities. *Experiment 3* replicates *Experiment 2* with toddlers, aged 18–30 months old.

Experiment 1

Experiment 1 tested adults' baseline preferences for individual-level versus relational-level causal hypotheses for an ambiguous sequence of evidence.

Methods

Participants and Design Participants were *N*=34 adults recruited from Amazon's Mechanical Turk. They viewed a 3-minute video and then made a forced-choice response. Six additional participants were tested but excluded for spending fewer than three minutes on the video or for failing an attention check.

Stimuli & Procedure Participants watched a video in which an experimenter placed pairs of blocks of distinct shapes and colors on top of a machine, which sometimes played music. The experimenter placed a total of four pairs of blocks on the machine; and the first and third pairs caused the machine to activate and play music. At test, the experimenter asked the participant to "Choose the blocks that will make my machine play music." Participants could

select one of two equally likely pairs: a "same" pair comprised of two individual blocks that had appeared in both successful trials (individual cause hypothesis), or a "different" pair comprised of the other two blocks that had appeared in the successful pairs (relational hypothesis) (see Figure 2).

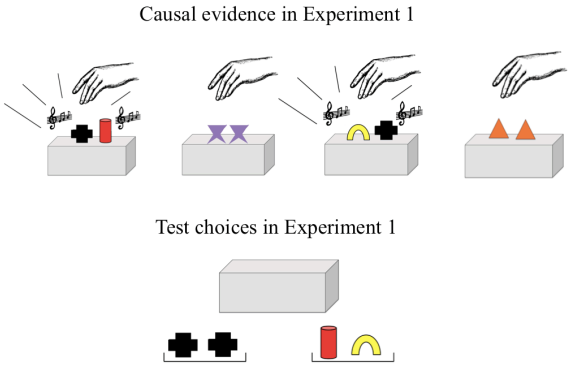

Figure 2: Stimuli for Experiment 1.

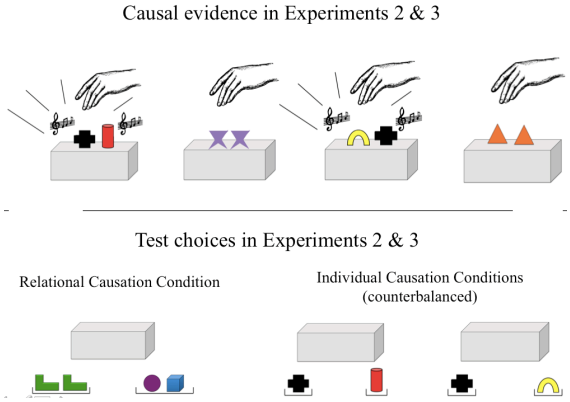

Figure 3: Stimuli used in both Experiment 2 (adults) and Experiment 3 (toddlers).

Results & Discussion Adults showed no baseline preference for either the individual level or the relational level hypothesis. Of the 34 participants, 19 chose the relational option (55.9%) and 15 chose the individual level hypothesis (44.1%). Results did not differ from chance, $t(33) = 0.68$, $p = .50$. These findings suggest that adult participants do not have a baseline preference for individual-level or relational-level causation for this ambiguous sequence of evidence that accords with both.

Experiment 2

Experiment 2 compared the responses of adult participants between two conditions. Both groups were trained on the same ambiguous evidence as *Experiment 1*, but each group was presented with a separate set of forced-choice options at test.

Methods

Participants and Design Participants were N=61 adults recruited from Amazon's Mechanical Turk. Participants were randomly assigned to view one of three, three-minute videos. Nineteen additional participants were tested but excluded due to spending fewer than the full three minutes on the video page or for failing an attention check question.

Stimuli & Procedure In each video, an experimenter presented the sequence of ambiguous causal evidence from Experiment 1. Then, the experimenter asked the participant to "choose the block(s) that will make my machine play music." At test, half of participants saw a video in which the experimenter presented a choice between novel *pairs*: a "same" pair (does not accord with the evidence) and a "different" pair (accords with the *relational hypothesis*). In the other two videos, participants received a forced choice between the block that was consistent with the *individual hypothesis* and one of the other two blocks that were part of the pairs in the successful trials. The decoy blocks were counterbalanced across two videos (see Figure 3).

Results & Discussion Adult participants performed significantly above chance in both conditions. In the individual causation condition (N=31), 74.2% of participants chose the appropriate individual block, $t(30) = 3.03$, $p = .005$. In the relational causation condition (N=30), 100% of participants chose the appropriate ("different") novel pair. Taken together, these results suggest that adults indeed hold multiple possibilities in mind as potential causal rules for explaining the ambiguous evidence they observe. Put differently, if participants were only considering one possible causal rule, then the results of *Experiment 2* would have resembled those of *Experiment 1*, with about half of the participants in each condition reliably tracking either one or the other of the two hypotheses. Instead, participants perform well above chance in both conditions.

Experiment 3

Experiment 3 replicated the results of *Experiment 2* with toddlers aged 18–30 months.

Methods

Participants and Design Participants were N=30 toddlers (M_{age} = 23.67 months) recruited from children's museums in southern California or from a lab database. Children were tested either in a quiet area of the museum or in a testing room in lab. Fourteen additional children were excluded from the sample due to experimenter error (4), failure to respond (8), failure to complete the task (1), or parental interference (1). Of the 30 children included in the sample, N=15 were randomly assigned to the relational causation condition, and N=15 were assigned to the individual causation condition.

Stimuli & Procedure Toddlers observed the same stimuli as the adults in *Experiment 2*. The only difference was that children observed this evidence performed in real time, rather than in a video. At test, the experimenter asked, "Can you make my machine play music?" The experimenter then pushed two trays containing either the two pairs of blocks (relational causation condition) or the two individual blocks (individual causation condition) towards the child. Whichever tray the child touched first was coded as the child's response.

Results & Discussion Toddlers performed significantly above chance in both conditions. Of the N=15 children in the relational causation condition, 13 (86.7%) chose the appropriate pair significantly more often than chance, p = .007 (exact binomial). Similarly, in the individual causation condition (N=15), 12 (80%) chose the appropriate individual block, p = .035. Taken together, these results suggest that children, like adults, tracked both the individual-level *and* the relational-level hypotheses in this ambiguous task.

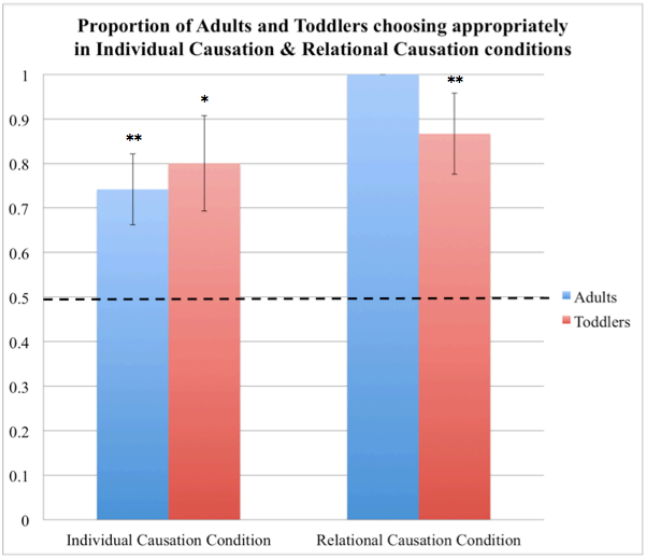

Figure 4: Results of *Experiments 2* (adults) & *3* (toddlers)

General Discussion

The results of these experiments represent initial evidence that both adults and very young children (aged 18–30 months old) can simultaneously track multiple, equally likely possibilities in a causal learning task.

First, *Experiment 1* established that adult participants consider a relational cause ("a pair of different blocks makes the machine play music") and an individual cause ("this block makes the machine play music") at roughly equal rates for a sequence of ambiguous evidence that was consistent with both hypotheses, suggesting no specific preference for one hypothesis over the other when no correct response was available. *Experiments 2 & 3* found that both adults and toddlers who observed the ambiguous evidence could flexibly make *either* the appropriate choice indicated by the relational rule *or* by the individual rule, depending on the test options available.

Taken together, the results of the latter two experiments suggest that participants were not simply attending to one hypothesis or the other, but rather considering both possibilities simultaneously. While it is possible that adult participants in *Experiment 2* were able to review or "replay" the evidence sequence they had seen in memory to select the appropriate response from the choices presented, it seems highly unlikely that toddlers, with their limited working memory capacity, would be able to do so (Alp, 1994; Garon, Bryson, & Smith, 2008). The age of participants in *Experiment 3* thus provides even stronger support for the notion that these very young learners attended to multiple hypotheses *simultaneously*.

To our knowledge, this is the first study to directly assess whether participants hold multiple causal hypotheses in mind as they reason from evidence. Furthermore, the causal possibilities that toddlers and adults must generate and evaluate in this task are relatively complex. Considering multiple causal hypotheses that can account for the evidence in this ambiguous scenario means that participants must consider the same entity both in the role of *individual cause* as well as of part of a pair comprising a *relational cause*. Although existing developmental research suggests that attending to individual entities and abstract relations are in tension, here toddlers seem to easily consider more than one type of property at once.

The results of *Experiment 3* are especially striking, given that research with older children has found that preschoolers and even early elementary school-aged children are unable to consider more than one distinct, equally likely possibility at a time (e.g., Beck et al., 2006; Redshaw & Suddendorf, 2016; Robinson et al., 2006). As such, the present findings raise important questions regarding the ontogeny of logical and counterfactual reasoning. In particular, these results suggest that the ability to track multiple possibilities in causal reasoning—specifically, the ability to track multiple causal rules or hypotheses—far precedes the ability to verbally predict or behaviorally prepare for multiple, equally probable possibilities. Future work should explore whether and how these different modes of reasoning about multiple hypotheses may be related.

These results may also carry important implications for understanding the development of epistemic uncertainty more generally. Under what circumstances do children come to appreciate that situations can have alternative possible outcomes? The same literature that reports that children cannot consider more than one possibility at once refers to the phenomenon of "premature closure," or children's general tendency to treat a determinate state of affairs as determinate (Robinson et al., 2006). This concept applies not only to counterfactual thinking but to children's understanding of language about epistemic modality—i.e., what is possible or necessary given the available evidence (Ozturk & Papafragou, 2015). By contrast, the present findings suggest that even very young children are

able to maintain "openness" about the different possible causal explanations that may account for the evidence they observe. Is there a fundamental difference between thinking about causal rules that *account* for outcomes, versus predicting those outcomes? Perhaps the former ability may give rise to the latter.

Although the results from these initial studies are certainly suggestive, additional work is needed. In particular, ongoing follow-up studies aim to assess whether individual toddlers can flexibly switch from one hypothesis to another, when presented with sequential choices in a within-subjects design. If the expected results obtain—that is, if 18–30 month olds demonstrate the kind of "hypothesis holding" that the current results suggest—this will strengthen the claim that even very young children are indeed tracking multiple possibilities simultaneously.

In sum, the present results demonstrate that even very young children can generate and simultaneously consider multiple, equally probable possibilities in a causal learning task. They appear to flexibly consider these multiple hypotheses, even when they are of very different *kinds*—i.e., individual vs. relational. These results therefore present a first step towards challenging existing claims that the ability to consider multiple possibilities is late developing, and open the door for future research.

Acknowledgments

Thanks to Alison Gopnik for her early contributions to the study design. We are grateful to members of the Early Learning & Cognition lab at UCSD, and particularly Sarah Gan, Alicia Lundhardi, Andrea Nishimi, Nicky Sullivan, and Emily To for their contributions to data collection. Thanks also to the Fleet Science Center, the San Diego Children's Discovery Museum, and the preschools, parents, and children who made this research possible.

References

Alp, I. E. (1994). Measuring the size of working memory in very young children: The imitation sorting task. *International Journal of Behavioral Development*, *17*(1), 125-141. doi: 10.1177/016502549401700108

Beck, S. R., Robinson, E. J., Carroll, D. J., & Apperly, I. A. (2006). Children's thinking about counterfactuals and future hypotheticals as possibilities. *Child Development*, *77*(2), 413-426. doi: 10.1111/j.1467-8624.2006.00879.x

Bonawitz, E. B., & Griffiths, T. L. (2010, January). Deconfounding hypothesis generation and evaluation in Bayesian models. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 32, No. 32).

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

Christie, S., & Gentner, D. (2014). Language helps children succeed on a classic analogy task. *Cognitive Science*, *38*(2), 383-397. doi: 10.1111/cogs.12099

Garon, N., Bryson, S. E., & Smith, I. M. (2008). Executive function in preschoolers: A review using an integrative framework. *Psychological Bulletin, 134*(1), 31-60. doi: 10.1037/0033-2909.134.1.31

Gopnik, A., & Sobel, D. M. (2000). Detecting blickets: How young children use information about novel causal powers in categorization and induction. *Child development*, *71*(5), 1205-1222. doi: 10.1111/1467-8624.00224

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. doi: 10.1037//0012-1649.37.5.620

Gopnik, A., Glymour, C., Sobel, D. M., Schulz, L. E., Kushnir, T., & Danks, D. (2004). A theory of causal learning in children: causal maps and Bayes nets. *Psychological review*, *111*(1), 3. doi: 10.1037/0033-295X.111.1.3

Gweon, H., & Schulz, L. (2011). 16-month-olds rationally infer causes of failed actions. *Science*, *332*(6037), 1524-1524. doi: 0.1126/science.1204493

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. doi: 10.1037/a0027440

Ozturk, O., & Papafragou, A. (2015). The acquisition of epistemic modality: From semantic meaning to pragmatic interpretation. *Language Learning and Development*, 11, 191–214. doi: 10.1080/15475441.2014.905169

Rafetseder, E., Cristi-Vargas, R., & Perner, J. (2010). Counterfactual reasoning: Developing a sense of "nearest possible world". *Child development*, *81*(1), 376-389. doi: 10.1111/j.1467-8624.2009.01401.x

Redshaw, J., & Suddendorf, T. (2016). Children's and apes' preparatory responses to two mutually exclusive possibilities. *Current Biology*, *26*(13), 1758-1762. doi: 10.1016/j.cub.2016.04.062

Robinson, E. J., Rowley, M. G., Beck, S. R., Carroll, D. J., & Apperly, I. A. (2006). Children's sensitivity to their own relative ignorance: Handling of possibilities under epistemic and physical uncertainty. *Child Development*, *77*(6), 1642-1655. doi: 10.1111/j.1467-8624.2006.00964.x

Walker, C. M., & Gopnik, A. (2014). Toddlers infer higher-order relational principles in causal learning. *Psychological science*, *25*(1), 161-169. doi: 10.1177/0956797613502983

Walker, C. M., Bridgers, S., & Gopnik, A. (2016). The early emergence and puzzling decline of relational reasoning: Effects of knowledge and search on inferring abstract concepts. *Cognition*, *156*, 30-40. doi: 10.1016/j.cognition.2016.07.008

Walker, C.M. & Gopnik, A. (2017). Discriminating relational and perceptual judgments: Evidence from human toddlers. *Cognition, 166*, 23-37.