In Press at Child Development

Toddlers learn and flexibly apply multiple possibilities

Mariel K. Goddu\textsuperscript{1}, J. Nicholas Sullivan\textsuperscript{2}, & Caren M. Walker\textsuperscript{3}

\textsuperscript{1}University of California, Berkeley

\textsuperscript{2}Stanford University

\textsuperscript{3}University of California, San Diego
Abstract

The ability to consider multiple possibilities forms the basis for a wide variety of human-unique cognitive capacities. When does this skill develop? Previous studies have narrowly focused on children’s ability to prepare for incompatible future outcomes. Here, we investigate this capacity in a causal learning context. Adults ($N=109$) and 18–30-month-olds ($N=104$) observed evidence that was consistent with two hypotheses, each occupying a different level of abstraction (individual versus relational causation). Results suggest that adults and toddlers identified multiple candidate causes for an effect, held these possibilities in mind, and flexibly applied the appropriate hypothesis to inform subsequent inferences. These findings challenge previous suggestions that the ability to consider multiple alternatives does not emerge until much later in development.

Keywords: cognitive development; causal inference; modal reasoning; learning; cognitive flexibility; relational reasoning
Much of adult mental life is devoted to *modal cognition*: representing and reasoning about possibilities (e.g., Phillips, Morris, & Cushman, 2019; Shtulman & Phillips, 2018). Whenever we consider something that *could, could not* or *might* be the case, given a particular set of conditions, we are thinking about possibilities—non-actual states of affairs\(^1\) (Kratzer, 2012). We may consider, for example, what we *could* eat for dinner, given what’s in the fridge (“stir fry”); what we *should* eat, given that we want to be healthy (“steamed vegetables”); or what we *might* eat, given that cooking is challenging and we may as well just order from our favorite burrito place.

Many aspects of mature human reasoning involve a particular subskill of modal cognition: the ability to reason about *multiple possibilities* simultaneously. When, for example, one’s growling stomach engenders the question, “What shall I eat?”, one often considers several possibilities at once: e.g., “I could cook or I could order a burrito.” When and in what contexts does the ability to simultaneously consider multiple possibilities develop?

On the one hand, this capacity has been prominently associated with a variety of human-unique and late emerging cognitive skills. These include future hypothetical thinking, or considering alternate futures (e.g., “How I’ll feel after eating a burrito” versus “How I’ll feel after eating steamed vegetables”), and counterfactual reasoning, or considering alternate pasts (e.g., “If I had eaten a bigger lunch, I wouldn’t be so tempted to order a burrito”; Beck & Riggs, 2013; Redshaw & Suddendorf, 2020, 2016; Seed & Dickerson, 2016; Suddendorf, 2006; Suddendorf, Bulley, & Miloyan, 2018; Suddendorf & Corballis, 2007; Tulving, 2005). Holding multiple possibilities in mind is also critical for logical thought, including contingency planning.

\(^1\) Technically, these are ‘mere’ possibilities, since actual states of affairs are also possible. For example: it is possible that you are reading this sentence right now—and actually, it just so happens that you are. But for the purpose of simplicity in this paper, we will use ‘possibilities’ to refer to mere possibilities.
(e.g., “I don’t remember which are better, black beans or pinto… So I’d better cover my bases and order both”) and reasoning with the disjunctive syllogism (e.g., “My favorite salsa last time was spicy, but I’m not sure which one of these two it was… It wasn’t the habanero, so it must be the other one”; Beck et al., 2006; Leahy & Carey, 2020; Mody & Carey, 2016).

On the other hand, the capacity to consider multiple possibilities seems to lie at the heart of a basic and early emerging cognitive skill: causal reasoning. Having placed my order, I may head out to pick it up—only to discover that my car will not start. Possibilities spring immediately to mind: Is it the battery again? The alternator? Am I out of gas? One prominent theoretical position and empirical approach implies that very young children are capable of exactly this type of causal reasoning, which involves discriminating among multiple hypotheses for the same outcome (e.g., Gopnik & Wellman, 2012; Gopnik et al., 2001, 2004). For example, even 16-month-olds can infer whether a toy is not working because it is broken or because they lack the skills to operate it (Gweon & Schulz, 2011).

Converging evidence from empirical and computational accounts (e.g., rational constructivism) suggest that even very young learners have a “hypothesis space” populated by many possibilities with varying likelihoods (e.g., Gopnik & Sobel, 2000; Gopnik & Wellman, 2012; Gopnik et al., 2001, 2004; Gweon & Schulz, 2011; Meltzoff, Waismeyer, & Gopnik, 2012). On this view, the ability to track multiple possibilities must be early emerging, since causal learning requires that learners assess the probability of a variety of alternatives. In addition, hierarchical Bayesian models have suggested that causal learning occurs
Simultaneously at multiple levels of abstraction: the same data can license multiple inferences at once\(^2\) (Goodman, Ullman, & Tenenbaum, 2011; Schulz et al., 2008; Tenenbaum et al., 2011).

Is the ability to reason flexibly about multiple possibilities a relatively sophisticated and late emerging cognitive capacity? Or is it an early developing and fundamental competence? Below, we review findings from previous studies, which have focused on a specific subset of possibilities— incompatible future outcomes—from which researchers have inferred that children may be unable to hold multiple possibilities in mind at once. We then critically examine the paradigms used in these studies and their theoretical conclusions. Finally, we propose a broader definition of “simultaneously considering multiple possibilities” and outline a novel causal reasoning task for probing the early emergence of modal reasoning in very young learners.

Many findings suggest that children’s capacity to consider multiple possibilities develops slowly. When young children are faced with a scenario that involves preparing for equally likely, but incompatible outcomes, they tend to fixate on a single possibility (e.g., Beck et al., 2006; Ozturk & Papafragou, 2015; Rafetseder, Cristi-Vargas, & Perner, 2010; Redshaw & Suddendorf, 2016; Robinson et al., 2006). In one study, 5-year-olds were introduced to a house with three doors and two bins of blocks: e.g., one containing all black, and the other 50% green and 50% yellow. During training, children learned the following rules: if the block is black, the experimenter would push it through the first door; if the block is yellow, they would push it through the second door; and if the block is green, they would push it through the third door. At test, the experimenter randomly drew a block from each of the bins. Although children succeeded in placing a tray to catch the block when the experimenter drew from the uniform

---

\(^2\) For example, if I learn that you like chocolate bars, this single “data point” might license multiple, distinct inferences. I may infer that you will prefer chocolate ice cream to vanilla (a relatively concrete inference), but I also may infer that you will prefer sweet to salty snacks (more abstract).
black bin, they failed to use multiple trays to prepare for the equally likely outcomes when the experimenter drew from the underdetermined green/yellow bin (Robinson et al., 2006). In related work, 3- to 5-year-olds failed to prepare for the possibility that an object (a mouse) could emerge from either of two openings at the bottom of an inverted Y-shaped tube: children tended to place a single mat for catching the mouse, rather than two (Beck et al., 2006). In a non-verbal version of this task, both non-human primates and children extended only one hand to prepare to catch a target (grape or ball) dropped into the tube. Although children’s tendency to extend both hands improved over sequential trials, less than 50% of children below age 4 produced this response on the first trial. Furthermore, even children who successfully generated the strategy on a given trial often regressed to one-handed responses on subsequent trials (Redshaw & Suddendorf, 2016; Suddendorf, Watson, Bogaart, & Redshaw, 2020).

Researchers have made the strong claim that children in these experiments blindly fixate on a single possibility (e.g., a simulation of an object traveling down the right or left branch of the y-shaped tube), mistake their guess for an actuality, and thus fail to entertain alternatives (Leahy & Carey, 2020). Critically, however, these behavioral paradigms have imposed a narrow (and stringent) operationalization of the ability to consider multiple possibilities. First, these methods define “possibilities” as future possible outcomes, and competence is measured in terms of preparing for each of these events. Second, the possibilities presented are always incompatible (i.e., the target could emerge either from the left or from the right). Here, we aim to broaden the scope of this research.

In the present experiments, we measure participants’ ability to simultaneously infer multiple possible causes for the same set of outcomes and apply this knowledge to inform their subsequent behavior. Thus, we operationalize “multiple possibilities” as “multiple candidate
hypotheses,” or possible explanations for a single set of observations. In contrast to previous studies, these possibilities are not mutually exclusive. Instead, both could be true. However, in line with previous work, the candidate causal hypotheses are “undetermined,” or “prefactual” (Beck et al., 2006, p. 425). Just as participants in the Y-shaped tube tasks generate an action before they observe where the object will emerge, participants in the current study form an inference and select an intervention before they have the opportunity to test the truth-value of their hypotheses. If children in our study fixate on a single, undetermined possibility and foreclose on all others, they should not be able to learn or apply both hypotheses. If, on the other hand, they are able to flexibly switch between these hypotheses in the course of the same reasoning task, this would provide strong evidence that they are capable of considering more than one possibility at the same time.

In the present study, participants observed a sequence of evidence that was equally compatible with two causal rules. Then, they were provided with opportunities to apply one, the other, or both rules to inform a novel inference and make a causal intervention. To do this, we modified a paradigm used in previous research to examine the development of relational reasoning (Carstensen et al., 2019). In this paradigm, participants are presented with evidence that accords equally well with a relational causal rule—i.e., a pair of same or different blocks placed on top of a toy cause it to play music—and an individual causal rule, in which a particular block is causally potent (see Figure 1). At test, participants are prompted to choose an intervention to produce the effect. They are provided with a choice either between one of two novel relational pairs (i.e., same or different), or between two individual blocks that had been previously associated with the effect (i.e., one block that was associated with both activations, and another “decoy” block that was associated with only one).
Using this novel method, we ask: If learners are presented with evidence that supports two possibilities, might they simultaneously infer and track both? In four experiments, we demonstrate that adults and toddlers learn and hold more than one hypothesis in mind for the same set of observations, and flexibly apply the appropriate rule in their subsequent inferences. These findings provide the first evidence for early competence in considering multiple possibilities.

**Experiment 1a**

In Experiment 1a, we presented adults with ambiguous causal evidence and asked them to produce the effect at test (see Fig. 1). If adults infer and track both hypotheses, they should be significantly more likely to select the test choice that accords with the relevant rule, regardless of the type of test choices provided (individual or relational, depending upon condition). A preliminary experiment established that adults do not have a baseline preference for either hypothesis: When relational and individual options were pitted against one another in a forced choice, adults chose between them at similar rates (56% vs. 44%, respectively), $X^2(1, N = 24) = .47, p = .49, n.s.$ (see Supplementary Materials for details). These results replicate the findings of extant research testing this preference (Kroupin & Carey, 2021).

**Method**

**Participants**

Participants were 61 adults recruited from Amazon’s Mechanical Turk, all from the United States, and all with an approval rating of >95%. Nineteen additional participants were
tested, but excluded due to spending less than full time on the video page or for failing the attention check question.

Materials

Materials included wooden blocks of distinct shapes and colors and one “toy.” The toy was composed of an opaque, cardboard box (10” x 8” x 4”), containing a wireless doorbell. The experimenter could activate the doorbell surreptitiously by using a remote-control button.

Procedure

Participants were randomly assigned to view one of two versions of a three-minute video. Participants in the relational causation condition ($N = 30$) viewed one video, and participants in the individual causation condition ($N = 31$) viewed the other. In both versions, an experimenter placed four pairs of blocks of distinct shapes and colors on top of the “toy,” which sometimes played music. The experimenter said, “Today, I’m going to play with my toy. Sometimes, when I put blocks on my toy, it plays music. But other times, when I put blocks on my toy, it does not play music. We’re going to try out some different blocks to find out which blocks make my toy play music. Okay?” The experimenter then placed a total of four pairs of blocks on the toy. The first and third pairs were composed of two different blocks and caused the toy to activate (AB, DA), whereas the second and fourth pairs were composed of two identical (same) blocks and failed to activate the toy (CC, EE) (see Figure 1). Critically, this sequence of evidence supported both individual and relational causal hypotheses: Since block A appeared in both causal “different” pairs, the individual causation (“object A”) and relational causation (“different”) hypotheses are consistent with the evidence.

Next, the experimenter asked participants to select a causal intervention. They presented two pairs of test blocks and said, “Can you choose the blocks that will make my toy play
music?” The pairs that were presented varied by condition. In the *individual causation* condition, participants received a forced choice between the individual causal block (A) and a decoy block (i.e., another block that was associated with the effect; object B or D, counterbalanced; see Fig. 1). In the *relational causation* condition, participants received a forced choice between two novel pairs of “same” and “different” blocks. If participants are able to infer both hypotheses, they should be able to select the test choice that accords with the correct rule in both conditions.

After watching their assigned video, all participants were prompted to advance to the next page, where they completed an attention check question that asked them to select which of two blocks had not appeared in the video.

**Results**

Results of Experiment 1a suggest that adult participants do infer multiple causal rules. In the *relational causation* condition ($N = 30$), 100% of participants chose the appropriate (“different”) novel pair. In the *individual causation* condition ($N = 31$), $74\%$ [95% CI: 55% to 88%] of participants chose the appropriate individual block, $p = .01$ (exact binomial), $d = .50$. That participants were slightly less successful in the *individual causation* conditions is sensible, given that the test choices involved two objects that had each been observed to be definitively associated with the effect, thus presenting a more challenging choice than the entirely novel relational pairs for which no concrete evidence had yet been observed. Additionally, learning the individual-level hypothesis entails tracking and recalling which of multiple candidate objects was most likely, whereas the relational hypothesis entails tracking only one relation—i.e., the “blessing of abstraction.” These results suggest that adults inferred both possibilities as potential
causal rules for explaining the ambiguous evidence they observed. They also provide a baseline for comparison with the performance of toddlers on the same task in Experiment 1b.

**Experiment 1b**

Experiment 1b used an identical design to investigate whether toddlers, aged 18–30 months, are also able to infer and track multiple causal hypotheses simultaneously from the same observations and use them generate an intervention at test. Toddlers were tested in person, rather than online.

**Method**

**Participants**

Participants were 56 toddlers ($M = 23.89$ months, range = 18.5–30.5, 30 females) recruited from children’s museums in a metropolitan area or from a lab database. Twenty-two additional children were excluded$^3$ from the sample due to experimenter error (6), failure to complete the task (15), or parental interference (1). Of the 56 children included in the sample, 28 were randomly assigned to the *relational causation condition* ($M = 24.07$ months), and 28 were assigned to the *individual causation condition* ($M = 23.71$ months). The sample size was chosen based on two previous studies that both used a causal relational reasoning paradigm that was highly similar to the one in the present task, and each of which found an effect size of 0.7 for 18–30-month-olds’ performance (Walker & Gopnik, 2014; Walker, Bridgers, & Gopnik, 2016).

**Materials**

Participants observed the same stimuli as the adults in Experiment 1a (Fig. 1). The only difference was that children observed this evidence performed in real time, rather than in a video.

**Procedure**

$^3$ All exclusions were participants tested in a museum setting, which proved distracting for this task in this age group.
Children were tested either in a quiet area of the museum or in a testing room in lab. After presenting the evidence, the experimenter said, “Now that you know how my toy works, I need your help finding the things that will make my toy play music! I have two choices.” The experimenter then presented two trays with the test choice blocks on them, one on each side of the table, each an equal distance from the child. The placement of the correct choice was counterbalanced between participants. The experimenter asked, “Can you point to the tray that has the things that will make my toy 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. The child’s response was coded as their first point or reach. If the participant did not respond, the experimenter prompted them by pointing to the trays in turn and saying, “These? Or these?” Videos of children’s responses were reliability coded by a second coder who was blind to the predictions of the experiment. The two coders agreed 100% of the time.

**Results**

Toddlers, like adults, were successful in both conditions: 79% of participants (95% CI: 63% to 94%) in the relational causation condition chose the appropriate pair, $p = .004$ (exact binomial), $d = .82$, and 75% of participants (95% CI: 59% to 91%) in the individual causation condition chose the appropriate individual block, $p = .01$, $d = 0.69$. Participants who answered correctly ($n = 43$, $M_{age} = 23.42$ months) were significantly younger than participants who answered incorrectly ($n = 13$; $M_{age} = 25.46$ months), $t(54) = -2.08$, $p = .05$, $d = .61$. This initially surprising age effect is consistent with prior work suggesting that older children learn an object bias that appears later in development (see Walker, Bridgers, & Gopnik, 2016.) Overall, toddlers’ performance was not different from the adults’ performance in Experiment 1a, $X^2(1, N$
These results therefore suggest that, like adults, 18–30-month-olds simultaneously inferred and tracked both causal hypotheses at two levels of abstraction and could apply them to produce an effect.

Experiment 2a

Results of Experiments 1a and 1b strongly suggest that adult and toddler participants are able to learn and apply multiple causal hypotheses for the same set of observations. However, evidence that individual learners maintain and flexibly apply both possibilities would provide converging evidence and further strengthen these findings. In Experiments 2a and 2b, we replicated these initial experiments as a within-subjects design with new samples of adults (2a) and toddlers (2b). To do so, we added a second test question to assess whether participants would make correct intervention selections in line with both the individual and relational causal rules on sequential trials.

Method

Participants

Participants were 48 adults recruited from Amazon’s Mechanical Turk, all from the United States, and all with an approval rating of >95%. Twenty-nine additional participants were tested, but excluded due to spending less than the full time on the video page or for failing the attention check question.

Materials

Materials were identical to those used in Experiment 1a.

Procedure

Experiment 2a used the same procedure as Experiment 1a, with one exception: after being asked to make an initial choice about which block(s) to put on the toy to make it play
music (either relational pairs or individual blocks), participants were then immediately prompted (without receiving feedback on their first response) to provide a response to the second question. The sequence of the test questions was counterbalanced across participants.

**Results**

This within-subjects experiment demonstrated that adults tracked both hypotheses: 71% [95% CI: 56% to 83%] of participants answered *both* the relational and individual questions appropriately, $p < .001$ (exact binomial), $d = 1.02$ (chance = 25%). Eight-five percent [95% CI: 72% to 94%] of participants chose appropriately on the first question, $p < .001$, $d = .80$, replicating the results of Experiment 1a, $X^2(1, N = 109) = .20$, $p = .66$, n.s. In addition, 81% of participants succeeded on the second question, $p < .001$, $d = .68$. There was no effect of question order, $X^2(1, N = 96) = .10$, $p = .75$, n.s.

**Experiment 2b**

Experiment 2b used an identical design to investigate whether toddlers, aged 18–30 months, are also able to learn and apply multiple causal hypotheses simultaneously. Again, toddlers were tested in person, rather than online.

Toddlers’ success in this task initially seems unlikely. First, limitations in early working memory capacity are well documented (e.g., Alp, 1994). Second, a large literature on the development of executive function—a set of regulatory processes including response inhibition and the ability to shift attention between tasks—suggests that even if toddlers can track both hypotheses, they may not be capable of applying this knowledge (e.g., Best & Miller, 2010; Garon, Bryson, & Smith, 2008). Given these challenges, success would provide strong evidence for early competence in the ability to consider multiple possibilities simultaneously.

**Method**
Participants

Participants were 48 toddlers (M = 23.98 months, range = 19.4–30.6, 27 males) recruited from a lab database and tested in a testing room in the lab. Seven additional children were excluded from the sample due to failure to complete the task (4), language comprehension issues (2), or parental interference (1). The sample size was again based on the two previous studies that each used a highly similar causal relational reasoning paradigm (Walker & Gopnik, 2014; Walker, Bridgers, & Gopnik, 2016).

Materials

Materials were identical to those used in Experiment 1b.

Procedure

Experiment 2b followed the same procedure as Experiment 2a, except that all procedures were presented live and in person, in the lab. As in Experiment 2a, all participants answered both a relational causation question and an individual causation question at test. The order of the questions was counterbalanced. After providing a response to the first question, both trays were removed (without feedback) and two new trays were presented to the child. The experimenter again asked, “Which tray has the thing(s) that will make my toy play music?” The trays were pushed towards the child, and again the child’s first point or reach was coded as their response. Videos of children’s responses were reliability coded, with 100% agreement.

Results

This within-subject experiment demonstrated that toddlers (n = 48), like adults, tracked both hypotheses: 50% [95% CI: 35% to 65%] of participants answered both the relational and individual questions appropriately, p < .001 (exact binomial), d = .53 (chance = 25%). The results of Experiment 2b replicate the findings from Experiment 1b: 67% [95% CI: 52% to 80%]
of participants chose the appropriate tray on the first question, \( p = .03, d = 0.35 \), with no difference between Experiments 1b and 2b, \( X^2 (1, N = 104) = 1.19, p = .28, n.s. \). Of the \( n = 46 \) children who provided an answer on the second question, 67% [95% CI: 52% to 80%] answered appropriately, \( p = .03, d = .34 \). There was no effect of question order, \( X^2 (1, N = 46) = .09, p = .76, n.s. \). There was no significant difference between the average age of participants who answered both questions correctly (\( n = 23; M = 23.57 \)) versus those who did not (\( n = 23; M = 24.87 \)), \( t(44) = -1.27, p = .21, n.s. \). The results of Experiment 2b therefore replicate and extend Experiment 1b to demonstrate that individual toddlers track multiple possibilities and can flexibly switch between an individual and relational hypothesis to select the appropriate intervention across contexts.

**General Discussion**

Together, these results suggest that children as young as 18–30 months can infer two possible causal explanations for the same set of observations, hold both in mind, and flexibly apply them based on the context. Given findings from prior work, researchers have suggested that preschoolers (and even school-aged children) may fixate on a single, undetermined or prefactual possibility and neglect alternatives (e.g., Beck et al., 2006; Leahy & Carey, 2020). The present findings challenge the generality of this proposal. Experiments 1a and 1b found that adults and toddlers can simultaneously learn and subsequently apply either an individual or relational causal rule, depending upon which is most appropriate for generating a successful intervention at test. Experiments 2a and 2b found that adults and toddlers can also flexibly switch between the two hypotheses to inform their selection. These results strongly suggest that participants inferred multiple possible causal rules simultaneously, and that they held both of these possibilities in mind.
The findings of Experiments 1b and 2b are particularly significant given previous empirical work and associated theoretical proposals. Prior studies have found that children are not able to behaviorally prepare for multiple, incompatible future outcomes (e.g., Beck et al., 2006; Leahy & Carey, 2020; Redshaw & Suddendorf, 2016). The fact that children are able to infer and track more than one undetermined causal rule, yet struggle to prepare for more than one undetermined physical outcome, opens an interesting set of developmental questions. Is the type of modal reasoning involved in causal learning more intuitive than—or perhaps a developmental precursor to—the later-emerging ability to “cover one’s bases” in light of physical uncertainty (Robinson et al., 2006)?

The present findings also suggest that reasoning about undetermined, incompatible possibilities may be a significant source of difficulty for the children in previous studies. In one sense, the current proposal is consistent with findings from studies using parallel tubes, in which children are able to prepare for multiple, compatible outcomes. Notably, however, those outcomes are more appropriately characterized as “inevitabilities”—non-actual events that are essentially pre-determined, and will necessarily become actual—rather than undetermined possibilities (Suddendorf, Crimston, & Redshaw, 2017). That reasoning about “exclusive-or” may be difficult for children is also consistent with recent findings that children do not reliably reason on the basis of the disjunctive syllogism (“A or B; not A; therefore B”) until age 5 (Mody & Carey, 2016). The difficulty of integrating multiple, incompatible future events may also be related to a more general developmental difficulty with integrating conflicting states within a single representation—e.g., in visual perspective-taking, appearance-reality, dual naming tasks, and Piagetian conservation tasks (Flavell, 1963; Suddendorf, 2006; Tomasello, 2019). Future studies investigating children’s ability to reason about incompatible possibilities in various
contexts may be informative for interpreting the link between the current results and these previous studies.

The present paradigm also relies on a less stringent definition of “simultaneity” than the one used in previous studies. Our results suggest that by requiring children to produce a specific and demanding behavioral response—contingency planning, or “covering one’s bases”—prior paradigms may have failed to detect children’s early developing competence in entertaining multiple, distinct alternatives. Here, we provide evidence that children are able to learn and hold more than one possibility in mind simultaneously—in the same reasoning task—without fixating on one or the other. The fact that participants in the current task responded to test questions sequentially may have reduced task demands, while nonetheless still demonstrating that they are able to simultaneously infer and actively maintain both possibilities in mind. Thus, in addition to their relevance for understanding the development of modal reasoning, these results offer an empirical demonstration of computational accounts claiming that knowledge is acquired at multiple levels of abstraction simultaneously (Goodman, Ullman, & Tenenbaum, 2011; Schulz et al., 2008; Tenenbaum et al., 2011). They also add to the growing literature demonstrating that young children are capable of reasoning about abstract relations, even when competing object-based hypotheses are available (Carstensen et al., 2019; Christie & Gentner, 2014, 2010; Walker & Gopnik, 2014, 2017; Walker, Bridgers, & Gopnik, 2016; Walker, Rett, & Bonawitz, 2020).

In sum, the current findings contribute to our understanding of the origins of human reasoning. Although the ability to track multiple, equally likely possibilities is critical for the development of a suite of higher-order cognitive skills, there has been little evidence for its early appearance. These results suggest that the ability to entertain distinct possibilities may emerge much earlier than previously believed, in the context of causal reasoning.
References


physical uncertainty. *Child development*, 77(6), 1642–1655. doi:
https://doi.org/10.1111/j.1467-8624.2006.00964.x

evidence: Abstract laws and preschoolers’ responses to anomalous

https://doi.org/10.1016/j.cub.2016.05.022


travel, and is it unique to humans?. *Behavioral and brain sciences*, 30(3), 299–313.
https://doi.org/10.1017/S0140525X07001975

determined, mutually exclusive possibilities in chimpanzees and children. *Biology
letters*, 13(6), 20170170.

Suddendorf, T., Watson, K., Bogaart, M., & Redshaw, J. (2020). Preparation for certain and
uncertain future outcomes in young children and three species of monkey. *Developmental
Psychobiology*, 62(2), 191-201.


Fig. 1. Schematic of causal evidence and test choices in Experiments 1a, 1b, 2a, and 2b.

Participants observed an ambiguous sequence of evidence that supports both the relational and individual causal hypotheses. In Experiments 1a & 1b, adults (1a) and 18–30-month-olds (1b) chose an intervention either between two novel pairs (“different” FG vs. “same” HH) in the relational causation condition, or between the repeated block (A) and a second, non-repeated decoy block (B or D, counterbalanced) that was also associated with the effect. Experiments 2a and 2b used a within-subjects design in which new samples of adults and toddlers made sequential choices between the relational test pairs and the individual block options, order counterbalanced.
Supplementary Materials

**Preliminary Adult Experiment**

This preliminary experiment assessed whether adults express a preference for either an individual cause (e.g., “the black block makes the machine play music”) or a relational cause (i.e., “different blocks make the machine play music”) (see Fig. 1S). This allowed us to establish a baseline rate of responding for each hypothesis: If adults are equally likely to infer either an individual or relational cause, we can conclude that neither is privileged, a priori.

**Method**

**Participants**

Participants included 34 adults recruited from Amazon’s Mechanical Turk, all from the United States, and all with an approval rating of >95%. Six additional participants were tested, but excluded, either for failing to watch the entire video of the procedure (i.e., spending less than three minutes on the test page), or for failing an attention check (see details below).

**Materials**

Materials included wooden blocks of distinct shapes and colors and one “toy.” The toy was composed of an opaque, cardboard box (10” x 8” x 4”), containing a wireless doorbell. The experimenter could activate the doorbell surreptitiously by using a remote-control button.

**Procedure**

Participants watched a three-minute video in which an experimenter placed four pairs of blocks of distinct shapes and colors on top of the “toy,” which sometimes played music. The experimenter said, “Today, I’m going to play with my toy. Sometimes, when I put blocks on my toy, it plays music. But other times, when I put blocks on my toy, it does not play music. We’re going to try out some different blocks to find out which blocks make my toy play music. Okay?”
The experimenter then placed a total of four pairs of blocks on the toy. The first and third pairs were composed of two different blocks and caused the toy to activate, whereas the inert second and fourth pairs were composed of two identical (same) blocks (see Figure 1). Next, the experimenter presented the two pairs of test blocks and said, “Can you choose the blocks that will make my toy play music?” Participants could select one of two pairs in a forced choice: a “same” pair comprised of two individual blocks that had appeared in both successful trials (individual causation hypothesis), or a “different” pair comprised of the other two blocks that had appeared in the successful pairs (relational causation hypothesis). Following their selection, participants advanced to a final page, which had an attention check question (“Which of these two blocks did not appear in the video you just saw?”)

Results

Results demonstrate that adults have no prior preference for either the individual or relational cause: 56% (95% CI: 39% to 73%) chose the test pair consistent with the relational cause and 44% (95% CI: 26% to 62%) chose the test pair consistent with the individual cause, $X^2(1, N = 24) = .47, p = .49, n.s$. These results suggest that adults lack a baseline preference, and replicate the findings of extant research testing this preference (Kroupin & Carey, 2021).
Participants observed four pairs of blocks placed on a toy. Two pairs activated the toy (AB and AD) and two pairs did not (CC and EE; trials were interleaved). This evidence is ambiguous, according equally well with an individual hypothesis, in which the single, repeated block (test choice AA) causes the effect, and a relational hypothesis, in which two “different” blocks make the toy activate (test choice BD).
Data availability

The data for these studies have been made public on Open Science Framework at the following link: [BLINDED].