Published in final edited form as:
Cogn Psychol. 2022 June ; 135: 101473. doi:10.1016/j.cogpsych.2022.101473.

The development of reasoning by exclusion in infancy

Roman Feiman[*,a], **Shilpa Mody**[b], **Susan Carey**[b]

[a]Department of Cognitive, Linguistic, and Psychological Sciences, Brown University

[b]Department of Psychology, Harvard University

Abstract

How do humans develop the capacity to reason? In five studies, we examined infants' emerging ability to make exclusion inferences using negation, as in the disjunctive syllogism (*P or Q; not P; therefore Q*). Inspired by studies of non-human animals and older children, Experiments 1–3 used an exclusion task adapted from Call's (2004) 2-cup exclusion paradigm and Experiments 4–5 used an exclusion task adapted from the blicket detector paradigm (Sobel & Kirkham, 2006). In both tasks, we found failure to make exclusion inferences at 15 months, fragile success at 17 months, and robust success by 20 months of age. These data converge with some prior evidence that fails to find a capacity to represent negation in infants younger than 15 months of age and conflict with other evidence from different paradigms that suggests infants do have this capacity. We discuss three different resolutions of these conflicting data, and suggest lines of further work that might adjudicate among them.

Keywords

disjunctive syllogism; mutual exclusivity; causal reasoning; reasoning by exclusion; negation; contrariety; logic

1. Introduction

Human thought supports logical inference, much of which is the stuff of everyday life. Walking into a restaurant, you might see a menu with two choices of taco: chicken or beef. If the server says, "We're out of beef," then you know (without having to ask) that your remaining option is chicken. You could do exactly the same for indefinitely many other pairs of taco fillings -- as well as movies, vacations, and so on -- which suggests that your reasoning process does not depend on the particular contents of the thoughts, but rather on their logical form: *P or Q; not P; therefore Q.* This form of inference is called the disjunctive syllogism, also known to non-logicians as reasoning by exclusion.

---

[*]Corresponding author. Address: Room 241, 190 Thayer St., Providence, RI, USA, 02906. roman_feiman@brown.edu.

For centuries, philosophers and psychologists have debated whether the capacity for logical inference is uniquely human and, if so, whether it awaits language development during ontogenesis (e.g. Bermúdez, 2003; Carruthers, 2002; Clark, 2006; Descartes, 1649; Davidson, 1999; Fodor, 1975; Penn, Holyoak, and Povinelli, 2008). Mastery of natural language, which allows understanding and expressing an unbounded number of inferences (complete with direct expressions of logical structure in words like “or” and “not”), has been the main diagnostic criterion of this capacity. More recent investigations, however, have aimed to develop non-linguistic diagnostics for assessing logical inference in creatures who lack language. While there has been progress in exploring non-human animals’ capacities for logically structured thought and inference (see Penn, Holyoak, and Povinelli, 2008; Völter & Call, 2017), there are fewer similar investigations of human infants prior to their mastery of the lexicon and compositional syntax of a natural language.

Here, we investigate reasoning by exclusion as a case study of the emergence of logically structured thought in infancy, building upon both the philosophical literature (Bermudez, 2003; Burge, 2010) and related studies of non-human animals (see Bohn, Call, & Völter, 2020; Völter & Call, 2017 for reviews). Reasoning by exclusion has two distinct logical components: disjunction (*or*) and negation (*not*). Negation excludes an option, *P*, from consideration by creating a representation, *not P*, that is *contrary* to it (meaning that *P* and *not P* cannot both be true).[1] Disjunction has a separate and dual function. It circumscribes the option space, forming a list of options to be considered; and it also yokes the options within that space to each other, such that thinking that *P* is false entails thinking that *Q* is true.

These different roles of negation and disjunction have different consequences in reasoning by exclusion. A reasoner able to represent *only* negation would be able to exclude an option (*P*) from consideration -- and would have the capacity to avoid choosing *P* -- but would not gain any information about the other option (*Q*) in the process. If only one option remains, then they would consider *Q* next as a matter of course, but without any increase in confidence that *Q* is true (effectively, as if *P* had not been checked already). On the other hand, a minimal consequence of representing a disjunction, *P or Q,* is that confidence in *Q* increases due to the elimination of *P.* Reasoning by exclusion can thus be broken down into two components: (1) exclusion on the basis of negation and (2) increased confidence in the remaining option on the basis of disjunction.

The emergence of both of these components of reasoning by exclusion has been the focus of much research, both in phylogenesis and ontogenesis. In studies of infants, there are three paradigms that have aimed to test for the ability to reason by exclusion prior to the ability to

---

[1] Indeed, a discrete compositional representation of negation is not actually required for this step; representing that any known fact (say, *R*) is *contrary* to one of the options is sufficient (e.g. that *the restaurant is out of beef* is contrary to *I can order beef,* because both cannot be true). A bit more formally, the difference between contrariety and negation is that while contrariety satisfies only the law of non-contradiction (LNC: both *P* and *contrary-P* cannot be true), negation satisfies LNC as well as the law of the excluded middle (LEM: either *P* or *not-P* must be true). For example, *Secretariat was a bird* is contrary to *Secretariat was a dog,* but neither proposition need be true (in fact, Secretariat was a horse). In contrast, if *Secretariat was a bird* is false, *Secretariat was not a bird* must be true (assuming that the presupposition that Secretariat exists is met, see Strawson, 1950; Horn, 1989 for discussion). Reasoning by exclusion relies only on LNC, not on LEM, so it can be supported either by contrariety or negation (see Horn, 2015). Thus, we do not differentiate between them in the present experiments and refer to the operation as “negation” for consistency with the existing empirical literature on reasoning by exclusion.

express its logical components with words like “or” and “not”: (1) Referent Disambiguation, (2) Object Identity Disambiguation, and (3) Object Location Disambiguation. Another paradigm, (4) Causal Disambiguation, has been used with older toddlers, and may be extended to infants. We argue that the first two paradigms have alternative explanations that do not require reasoning by exclusion, but that these alternatives do not apply to the last two paradigms.

1.1. Referent disambiguation.

In referent disambiguation (also called “mutual exclusivity”) studies, the child is shown two objects – one for which they know a label (e.g., a ball) and one novel (e.g., a vacuum tube) – and is told to find the “dax.” Many studies have found robust success (looking at or choosing the novel object) around 18 months of age (Bion, Borovsky, & Fernald, 2012; Byers-Heinlein & Werker, 2009; Halberda, 2003; White & Morgan, 2008; Yurovsky and Frank, 2017). Two of these studies also found failure at younger ages (Halberda, 2003; Yurovsky & Frank, 2017), but three others report success with infants as young as 12 and 14 months (Jin & Song, 2017; Pomiechowska, Brody, Csibra, & Gliga, 2021; Yin & Csibra, 2015). Many analyses of success propose that the child is excluding the object with a known label from consideration due to a “mutual exclusivity constraint.” This constraint specifies that a single object cannot have two labels. Because the novel label is *incompatible* with the familiar one, the child excludes that object as a candidate referent and chooses the other object (e.g., Markman & Wachtel, 1988).[2] However, an alternative explanation with significant support is that infants are *positively* drawn to apply novel names to novel objects, without considering and rejecting objects whose labels are known (Golinkoff, Hirsh-Pasek, Bailey, & Wenger, 1992; Mather & Plunkett, 2010, 2012; Merriman & Bowman, 1989; Mervis & Bertrand, 1994). If this strategy underlies infants’ success at referent disambiguation tasks, then success provides no evidence for reasoning by exclusion.

Current evidence is inconclusive about whether infants solve the referent disambiguation task by mapping novel labels to novel objects or by using one or both components of reasoning by exclusion. Halberda (2006) established an online processing signature of the exclusion component within the referent disambiguation paradigm in adults and 3- to 4-year-old children. Using eye tracking, he found that *only* when participants happened to be looking at the novel object at the time when the novel label was uttered, they performed a “double-check” -- their gaze moved away to the familiar object and then back to the novel object, remaining there -- consistent with their considering and excluding the familiar object as a possible match for the novel label. In contrast, Halberda (2003) found that 18-month-olds looked more to the novel object when hearing the novel label, but did not find any evidence of double-checks (personal communication). It is unknown at what age this signature of an exclusion inference emerges.

Bion, et al. (2012) established a separate signature of increased confidence in the remaining option (the disjunction component of reasoning by exclusion) within the same paradigm. They first presented infants at 18, 24, and 30-month olds with the standard referent

---

[2] This abstract characterization holds whether the mutual exclusivity constraint derives from lexical constraints or pragmatic reasoning.

disambiguation task, followed immediately by a test of retention -- the same novel label presented together with the same novel object now alongside a second, completely unfamiliar novel object. If, upon hearing the novel label in the first mutual exclusivity phase, children did not merely look at the novel object, but reasoned by exclusion to conclude that it is likely the label's referent (i.e. they gained increased confidence in the remaining option), they should have been able to re-identify the object at the second phase. But Bion, et al. found that prior to 30 months children looked equally between the two novel objects, suggesting that younger children may not have related the two alternatives in the preceding mutual exclusivity phase via disjunction.

Thus, while two lines of studies have found evidence for both components of reasoning by exclusion in referent disambiguation tasks in 3-year-olds and adults, no signature of either component has yet been observed in children under 2 ½ years of age. In younger children, the observed behaviors are consistent with positively matching a novel label to a novel object.

1.2. Object identity disambiguation.

Taking a different approach to the question of whether infants can reason by exclusion, Cesana-Arlotti, et al. (2018) developed a task that cannot be solved by matching novelty to novelty. In their task (Figure 1), infants watched a movie that started with two objects: e.g., a snake and a ball. In the crucial Inference condition, a screen rose up to cover both objects, then a cup appeared, swooped in behind the screen, and came out with the top of an object peeking out over its rim. The snake and ball were designed to look identical when only the top was visible, so a viewer could not tell which object was in the cup and so might set up a representation of a disjunction, *either the ball or the snake is in the cup*. Next, a snake briefly emerged from the side of the occluder before retreating behind it again. Representing that the snake was behind the occluder could be taken as *contrary* to the snake being in the cup, licensing the elimination of that option from the disjunction, and the inference that it is the ball that is in the cup. In a control No-Inference condition, there was no ambiguity about the locations of either the ball or the snake throughout the whole event.

Cesana-Arlotti, et al. used several sources of evidence to argue that infants reasoned by exclusion. First, infants looked longer at the final inconsistent events than the final consistent events in both conditions: either a ball emerging from behind the occluder (implying three objects in total), or a snake being revealed in the cup (implying two snakes and no ball). However, as Cesana-Arlotti and colleagues acknowledge, these results do not necessarily mean infants were reasoning by exclusion. Infants' longer looking could reflect their attempting to reconcile the initial representation of a snake and a ball, held in working memory, with the final evidence either for two snakes or for three total objects.

Stronger evidence for reasoning by exclusion comes from comparing patterns of eye movements and pupil dilation between the Inference and No-Inference conditions during the period in which an inference might have been made (the 'potential deduction phase'; see Figure 1). Just after the snake emerged from behind the occluder, infants' pupils dilated more in the Inference condition, suggesting greater attention or processing, and they shifted their gaze more from the occluder to the cup, consistent with their inferring its contents.

Furthermore, more pupil dilation and shifts to the cup in the inference condition predicted greater looking times on the final violation of expectancy test trials. Cesana-Arlotti and colleagues take these data to show both that infants eliminated the possibility that the snake was in the cup (the function of negation) and that the infants concluded that it is the ball that must therefore be in the cup (the function of disjunction). They found these signatures in 12- and 19-month olds, and 14-month-olds also succeeded in a variant that further requires integrating the representation that results from this inference with a representation of agents' preferences (Cesana-Arlotti, et al., 2020).

Although these findings provide strong evidence that the child is inferring the identity of the object in the cup in the inference condition, they do not specify the nature of this inference. There is an alternative explanation, on which the inference requires neither the negation nor disjunction components of reasoning by exclusion. Like other results that show infants can track the identities of objects even as they temporarily lose perceptual access to them, these findings can be explained in terms of infants' capacity to monitor the consistency of working memory models of a scene relative to subsequent percepts (for reviews, see Carey, 2009; Kibbe, 2015; Leslie, Xu, Tremoulet, and Scholl, 1998).

On this alternative explanation, infants create a mental model of the initial scene and hold it in working memory: e.g. one snake and one ball on the stage. As the scene unfolds, they simply build a perceptual representation of what they are looking at, monitoring for consistency between the unfolding scene and the initial model. When the objects are occluded and the cup scoops one up, the perceptual representation of what is in the cup now contains a 'bare' object file (a spatiotemporal index with no other properties; see Scholl, 2001), or perhaps an object with a red top part. This scene is still consistent with the initial model. At this point, infants need not represent any specific alternatives about what is in the cup or what is behind the screen. Next, they see the snake emerge from behind the screen, and they specify the current location of the snake that had been held in their working memory model. Now infants are in a position to make an inference. Finding out where the snake is leaves only two unknowns that just happen to match each other -- an object with an unknown location (the ball that was present at the start, a representation of which is still held in working memory), and a location (the cup) visibly containing an indexed object, which also happens to look like that ball. This allows for a 1-to-1 mapping of the spatiotemporal individual to the kind – that object in the cup is the ball! – producing a genuine inference, which may result in corresponding increases both in looks to the cup and in pupil dilation. Critically, however, on this account infants never wonder whether what is in the cup is the ball *or* the snake, and they never consider the possibility that the snake is in the cup (let alone exclude this possibility from consideration once they see the snake emerge from behind the screen). That is, they represent neither negation or disjunction. The appearance of the snake is only incidentally informative about the location of the ball because it happens to leave only one way to specify the two remaining unknowns (where the ball is, and what is in the cup). Hereafter, we refer to this leaner account as "1-to-1 mapping" (see also Jasbi, et al., 2018).

More generally, a 1-to-1 mapping process is an alternative to reasoning by exclusion whenever there are two entities, two properties, and the constraint that each of the entities must have no more than one of the properties. These conditions hold not only in Cesana-Arlotti, et al.'s paradigm, but in other tasks as well. In the referent disambiguation tasks described above, mapping novel names to novel labels is a kind of 1-to-1 mapping account, in which the mutual exclusivity constraint specifies the requisite condition that each object can have at most one label. So again for other tasks. Four- and 5-year-olds shown a stranger's face alongside a familiar one whose voice they know will associate a novel voice with the stranger (Moher, Feigenson, & Halberda, 2010), a result recently extended to 8-month-olds (Ekramnia, Mehler, & Dehaene-Lambertz, 2021). Assuming the constraint that each person has one voice (which must be assumed for reasoning by exclusion, too) such tasks can also be solved by a 1-to-1 mapping process.

Importantly, that there are possible leaner interpretations of infants' reasoning in these tasks does not establish that reasoning by exclusion is not involved. However, if infants can reason by exclusion as early as 12 months of age, they should also succeed on tasks to which these leaner accounts do not apply. In the present studies, we test infants on two such tasks: one task that involves disambiguating objects' locations, and another that requires disambiguating causal properties.

1.3. Object location disambiguation

The paradigm most used to explore reasoning by exclusion in nonhuman animals, introduced by Premack and Premack (1994; Premack, 1995) and developed further by Call (2004), requires disambiguating the location of a single object. In Call's (2004) version, chimpanzees, gorillas, and bonobos saw an experimenter hide a piece of food in one of two covered cups behind a screen. After the screen was removed, the apes were shown the inside of the empty cup, which licensed excluding it from consideration. The two cups were then placed within reach. The apes performed nearly perfectly in choosing the non-empty cup right from the first trial, with no evidence of learning. Subsequent studies have found similar success across a range of species: in at least some individual capuchin monkeys (Sabbatini & Visalberghi, 2008), African grey parrots (Pepperberg et al., 2013, 2018; Schloegl et al., 2012), and elephants (Plotnik, et al., 2014), among many others (see Völter & Call, 2017, for review). The task has also been adapted to preschool-age children, who succeed at 3 years (Hill, Collier-Baker, and Suddendorf, 2012), 23 months (Mody and Carey, 2016), and 20 months of age (Feiman, Mody & Carey, 2017).

While the 2-cup task can be solved by reasoning by exclusion, deploying both *not* and *or*, it provides better evidence for negation than disjunction. While it is possible that participants represent that the object is in one *or* the other location, it is also possible that they merely constrain their search space to the two visible containers. After excluding the empty cup, they may check the remaining option because it is the only one left, but without being certain that the object is there; indeed, as if they had not already seen that the other location is empty.

Consistent with this alternative, three converging lines of research have found the signature of disjunction in location disambiguation tasks -- increased confidence in the remaining option -- only in children at 4 years of age and older, but not in younger children or in non-human animals. First, Watson et al. (2001) showed dogs and 4- to 6-year-old children an object being hidden behind one of three screens. Participants were allowed to search, checking behind each screen in any order they chose. When they happened to search the two empty locations first, Watson, et al. measured how long participants took to move to the last remaining location. These authors reasoned that if participants related the options with a disjunction (the object is in location 1 *or* 2 *or* 3), they would speed up once they discovered two empty locations, reflecting increased confidence that the object must be in the third. This is what the 5-year-olds did. In contrast, dogs slowed down before the third search, as if an associative mapping between the object and the screens was being extinguished.

Call and Carpenter (2001) used a task with the same structure to test 2 ½-year-old children and chimpanzees, but looked for a different behavioral signature of increased confidence in the remaining option. An object was hidden in one of three containers, and they tested whether participants who had happened to check two empty containers first would simply reach into the third, without looking inside. Neither chimpanzees or 2 ½-year-olds reliably did, suggesting they were not confident that the object would be there. Indeed, they reached without looking on about 20% of trials whether they had just checked only one of the other tubes, or both. This result could not be due to a general unwillingness to reach without looking -- both groups did so significantly more often on trials in which they had seen directly which container the object was hidden in.

Finally, Mody and Carey (2016) sought evidence for increased certainty in the remaining option within a location disambiguation task in yet a third way. They tested children on a 4-cup task that effectively doubled Call's (2004) display. First, one sticker was hidden in one pair of cups behind one occluder and a second sticker was hidden in another pair of cups behind a separate occluder. Then, one of the 4 cups was shown to be empty, and children were invited to search for a sticker. If children represent that either that cup *or* its mate contains a sticker, they should choose the other cup within that pair (100% chance of finding a sticker) over either cup in the second pair (a 50/50 chance). Mody and Carey found that children under 4 years of age chose between the certain option and one of the cups on the uncertain side at chance, showing no signature of representing disjunction between the two cups of the first pair (see also Grigoroglou, et al., 2019, but see Gautam, et al., 2021, and Leahy, under review, for discussion). At the same time, they virtually never (3 out of 208 trials) looked in the empty cup, which shows that they had excluded it.

In sum, while children under 4 years of age show no signature of increased certainty in the remaining option in these location disambiguation tasks, by 20 months of age they successfully exclude an option -- they do not search in the location they saw was empty. These data suggest that while young children do not relate multiple locations via disjunction, they do use negation to rule individual locations out.

Note that the exclusion inferences in these tasks cannot be explained by the alternative 1-to-1 accounts of success in the mutual exclusivity paradigms or Cesana-Arlotti, et al.'s tasks. In Cesana-Arlotti, et al.'s task,, seeing positive information about what is in one location (the snake behind the screen) might only incidentally inform what is in the other location (in the cup) because it leaves only one way to create a 1-to-1 mapping between currently indexed spatiotemporal objects and the object-kinds represented in the working memory model of the initial scene. The location disambiguation tasks have no analog to this ‘incidentally informative’ step. In the 2-cup task, there are not two objects and two properties to be put into 1–1 correspondence; there is only one object, and two locations. The empty location can only be relevant to finding the object if *empty* (or *no object there*) is treated as contrary to *contains object*, and used to exclude that location from consideration.

The earliest age at which infants exclude an option in Call’s 2-cup task has not yet been established, although relevant data derive from another location disambiguation task: Piaget’s invisible displacement paradigm. All variants of invisible displacement that are relevant to reasoning by exclusion share a common structure (see Gopnik, 1984, for review). An experimenter first shows the infant an object. The experimenter then hides the object in their hand, and moves the hand behind an occluder (in some variants, behind multiple occluders in a row), and when the hand emerges again, infants are shown that it is now empty. Success on this task counts as searching behind the occluder(s) where the hand had been. Piaget himself did not construe the invisible displacement task as a test of reasoning by exclusion; rather he located the failures of younger infants as reflecting limitations on their representations of object permanence (Piaget, 1954). Although this interpretation has been rejected by a massive literature that shows infants do represent the persistence of objects out of their view (see Carey, 2009; Spelke & Kinzler, 2007; Stavans, et al. 2019, for reviews), success on this task does require excluding the empty hand as an option for where the object is and constraining search to the visited locations, similar to the 2-cup task and other location disambiguation paradigms. Infants succeed on the versions of invisible displacement that involve exclusion by 18 months of age (Gopnik, 1984), though failure at younger ages could be caused by insufficient executive function (Diamond, et al., 1997), rather than an inability to compute the exclusion inference.

Experiments 1–3 seek to establish the earliest age of success on Call’s 2-cup task. Doing so is important for three reasons. First, as mentioned above, this paradigm is not subject to the same alternative explanations as either referent disambiguation or identity disambiguation tasks (matching novelty-to-novelty or 1-to-1 mapping). Second, because this paradigm has been systematically used in studies of older children and non-human animals, the comparison enables a richer interpretation of data from infants and a more precise understanding of just what is being tested -- a signature of negation, in the service of an exclusion inference, but not of disjunction.

Finally, the 2-cup task can provide a meaningful signature not only of success, but also of failure to exclude an option: searching randomly between the two cups after having just seen that one of them is empty. If infants search in the empty cup half of the time, it would suggest they had learned nothing from that evidence about the location of the object. In contrast, if infants failed to follow the hiding event or to restrict the search space to the two containers where the toy could be, they might search elsewhere in the room, try to check if the experimenter is holding something, or ask for help; but they would not be predicted to ever search in the bucket they just saw to be empty, let alone to do so on half of the trials.

In principle, having constrained the search space to two containers, chance performance might be due to any performance demand that makes the task overwhelming or confusing.

However, there is independent evidence that the non-logical demands of the 2-cup task are well within the capacity of infants by 12 months of age. For example, in Feigenson and colleagues' bucket choice task, crackers are placed into two buckets one at a time (e.g., 2 crackers into the left bucket, 3 into the right) and infants can choose only one of the buckets. This requires establishing working memory models of two sets of 1–3 objects in two different containers, updating each model as additional objects are added, comparing the two final models, and being able to inhibit responses to the last-manipulated container (Feigenson, Carey & Hauser, 2002; Feigenson & Carey, 2005; see Zosh & Feigenson, 2009; Stavans, et al. 2019 for review). Ten and 12-month-old infants succeed at these tasks as long as there are three or fewer objects in each bucket. In contrast, the 2-cup task involves only one object hidden in one of two containers. Thus, the 2-cup task imposes vastly fewer working memory demands and some of the same inhibitory demands (being able to avoid choosing the last manipulated container) as the bucket choice tasks that even 10 month olds succeed at.

A more general requirement on performance limitation explanations of failure is that they must be specific to the tasks on which children fail, and cannot also apply to the tasks on which they succeed. Unlike the cracker choice task, the 2-cup task requires representing a single object, located ambiguously in one of two containers, as well as an exclusion step. Whatever additional demands this might impose, those demands are shared with Cesana-Arlotti, et al.'s tasks, on which 12- and 14-month-olds succeed, assuming the rich interpretation on which infants are reasoning by exclusion. This interpretation requires representing the ambiguity of possible locations for each of two objects, then disengaging from the excluded option and increasing the likelihood that the other object is the one hidden in the cup. Moreover, all of these computations together must take fractions of a second in order to happen during the "potential deduction phase". If the richer interpretation is correct and infants are reasoning by exclusion in Cesana-Arlotti's task, we would expect infants as young as 12-months of age to succeed at exclusion inferences in the 2-cup task, as 20- and 23-month-olds do (Feiman, et al, 2017, Mody & Carey, 2016).

1.4. Causal Disambiguation.

One other paradigm has previously aimed to test children's ability to reason by exclusion in a different context: reasoning about the causal properties of objects rather than their locations. Studies using the "blicket detector" paradigm have found that children as young as 2 years old are sophisticated causal reasoners (Gopnik, Sobel, Schulz, & Glymour, 2001; Waismeyer, Meltzoff, & Gopnik, 2015; Walker & Gopnik, 2014). In these studies, children interact with a toy that lights up when some, but not all, blocks are placed on it. Children are shown evidence about the causal status of individual blocks, and are then asked to make judgments about or choose among the blocks. Some kinds of trials (dubbed "indirect screening-off" in this literature) require children to make exclusion inferences (Sobel & Kirkham, 2006; Sobel, Tenenbaum, & Gopnik, 2004). For example, in Sobel and Kirkham's (2006) study, two blocks were placed on the toy together, which activated it. Next, one of those blocks was placed on the toy alone, which did not activate it. Twenty-four-month-olds excluded the inert block from consideration and attempted to activate the toy using the other block.

In Experiments 4 and 5, we search for the earliest age of exclusion inferences on these indirect screening-off trials. Like the 2-cup task, these trials are not subject to the leaner alternative explanations of either the referent disambiguation or identity disambiguation tasks. There is no 1-to-1 mapping process, no two objects with mutually exclusive causal affordances. Indeed, seeing one block fail to activate the machine leaves two options open – either both blocks are needed, or the other block is sufficient on its own. Taken together with the 2-cup task, these experiments provide another separate measure of infants' capacity to make exclusion inferences on the basis of negation.

2. Experiment 1. Two-bucket location disambiguation task.

Experiment 1 seeks the earliest age of spontaneous success on Call's 2-cup task, replacing the cups with larger buckets to allow hiding a toy inside. Mody and Carey (2016) found robust success with 23-month-olds (Cohen's $d = 1.3$) on this task, and Feiman et al. (2017) reported success in a closely related procedure at 20 months of age (Cohen's $d = 0.7$).[3] We therefore began with 17-month-olds and 14-month-olds.

2.1. Methods

2.1.1. Participants—Following the sample size in Mody and Carey (2016), we tested 17-month-olds ($N = 24$, $M_{age} = 17.70$ months, range = 17.1–18.5, 11 boys) and 14-month-olds ($N = 24$, $M_{age} = 14.4$ months, range = 13.6–15.4, 13 boys). An additional 11 14-month-olds and 13 17-month-olds were excluded for not searching on any of the warm-up trials (16), parental interference (3), completing fewer than three usable test trials (4), or experimenter error (1). In all studies reported here, participants were recruited by phone and email and were tested at the Laboratory for Developmental Studies at Harvard University. The subject pool consists of mostly upper middle class families, approximately 70% Caucasian, with the rest being Asian, Hispanic and African American. Children were given a small gift and parents were compensated $5.00 for travel expenses.

2.1.2. Materials & procedure.—The stimuli consisted of four pairs of cloth-lined buckets (approx. 1' high) and a large black screen (approx. 2.5' high and 6' across). Each trial used two identical buckets, placed 3' apart; the color of the buckets varied across trials to reduce perseveration. A ball was hidden in one of the buckets on each trial. Infants were held on their caregiver's lap, who sat on the floor 6' away from the experimenter. The experimenter's upper body was fully visible, and the toy was initially held at midline by both hands. The hands were lowered behind the and then visibly separated, so that the child could see that the experimenter's hands were both being lowered into the two buckets, but could not know which hand contained the toy. Caregivers were asked to close their eyes while the ball was being hidden and the empty bucket was revealed. Each child participated in two warm-up trials and four test trials.

2.1.2.1. Warm-up trials.: Each session started with two warm-up trials using only one bucket. On the first warm-up trial, the experimenter held the ball above the bucket, called for

---

[3] Feiman, et al.'s (2017) procedure used two different containers, a bucket and a truck, and also included affirmative trials in which infants saw the ball in the bucket. Because of these differences, we do not report analyses comparing those data to the present study.

the child's attention, and then lowered the ball into the bucket with both hands in full view. She then asked the child to find the ball. On the second warm-up trial, the experimenter placed the screen in front of the bucket, lowered the ball into it, then removed the screen and asked the child to find the ball. If children failed to search on the second warm-up trial, they were given another identical warm-up trial. To proceed to the test trials, children had to search in the bucket on at least one of the two or three warm-up trials.

2.1.2.2. Test trials.: There were 4 test trials. On each test trial, the experimenter placed two identical buckets in front of herself, 38" from each other. She placed the screen in front of the buckets and held the ball above the center of the screen. She caught the child's attention and lowered the toy with both hands, saying, "look where it's going!" When her hands were behind the screen, she separated them and lowered each hand into a bucket, secretly depositing the ball, then removed the screen.

The experimenter demonstrated that one bucket was empty by turning it upside down, shaking it, and showing the child the inside of the bucket, and then placed it back in its original position. She then said "Can you find the ball?", whereupon the parent released the child to go search. Children's responses were coded by which bucket they touched, looked into, or stood in front of first. If the child did not approach either bucket within 5 seconds, the experimenter encouraged them to search for the ball until they approached a bucket or approximately 10 seconds elapsed. If they did not find the ball themselves, the experimenter showed them where it was; regardless of their actions, at the end of the trial, children were given the ball. Two orders for the location of the toy were constructed – (left, right, right, left) and (right, left, left, right)– and each order was used for half the children. Figure 2 shows the procedure.

2.2. Results.

For all experiments, data and analysis code are available at https://osf.io/ma982/. For comparison, the present analyses include the previously published data from Mody and Carey's 23-month olds (N = 24, M_{age} = 23.6 months, range = 23.0–24.0, 13 boys). For one 17-month-old and six 14-month-olds, one trial was excluded because the caregiver released the child early (4) or the child failed to approach a bucket (3). Using the remaining test trials, we computed each child's percentage of correct searches. Results are summarized in Figure 3.

Seventeen-month-olds approached the correct bucket on 51% of trials, which did not differ from chance on a Wilcoxon signed-rank test, V = 70, $p = .935$. There was a marginally significant learning effect between the first two (42% correct) and last two (60%) test trials, V = 24, $p = .063$. Similarly, 14-month-olds searched in the correct bucket on 44.4% of trials, which did not differ from chance, V = 50.5, $p = .372$. At this age there was no evidence of learning between the first two (46% correct) and last two (42%) test trials, V = 32, $p = .666$. Infants were not simply confused or unresponsive -- they did search, constraining their search to the two buckets, but chose the bucket that was just shown to be empty half of the time. This is the signature of failing to exclude that bucket from the choice space.

We found no evidence that failure at exclusion was due to children perseverating in searching on the same side across all four trials: only four 14-month-olds and three 17-month-olds did so (binomial tests based on 1/8 chance rate; 14-month-olds: $p = 0.532$; 17-month-olds: $p = 1$). Furthermore, there was no indication of success on the first trial: in both age groups, 12 of 24 children approached the correct bucket, sign tests, both $p = 1$). In contrast, the 23-month-olds (from Mody & Carey, 2016) approached the correct bucket on 79% of trials, which was significantly greater than chance, V = 203.5, $p < .001$ and were marginally successful from the first trial (71% correct; sign test, $p = 0.064$), with no evidence of learning between the first two (77%) and last two (79%) trials (V = 30.5, $p = 0.85$). Combining the data from the 23-month olds with those from the 14- and 17-month-olds, a Kruskall-Wallis test found a main effect of age on infants' performance ($\chi^2(2) = 17.1$, $p < 0.001$). Post-hoc Mann-Whitney tests revealed that 23-month-olds chose the correct bucket more often than 17-month-olds (W= 438.5, $p = .001$) and 14-month-olds (W = 469, $p < .001$). The two younger age groups did not differ from each other (W = 332.5, $p = 0.35$).

Choosing the bucket that was just shown to be empty is consistent with children lacking the capacity for exclusion. However, the procedure of Experiment 1 could also underestimate infants' competence due to high executive function (EF) demands. On every trial, after the screen was removed, the experimenter drew participants' attention only to the empty bucket by manipulating it and showing the infant what was inside it. To go to the other bucket, infants would have to disengage attention and inhibit any impulse to search in the empty bucket to which attention was drawn. While the literature on infants' working memory capacity shows even 10-month-olds can track and remember an object hidden in one of two buckets, and can inhibit attention to the last manipulated bucket, the experimenter in those studies always manipulates both buckets in sequence before the choice (Feigenson & Carey, 2005; Feigenson, Carey & Hauser, 2002). This leaves open the possibility that drawing attention to only one of the buckets is enough to cause 17-month-olds to fail, especially given independent evidence that both endogenous disengagement of attention and response inhibition pose difficulties at these ages (Diamond, 2013). Experiment 2 seeks to ease these demands while keeping the reasoning structure of the task the same.

3. Experiment 2. Two-bucket location disambiguation with both containers manipulated

Experiment 2 tests whether children under 20 months can succeed on the test trials when attention is drawn to *both* containers prior to their choosing one, and further seeks the earliest age of success on Call's 2-cup task.

3.1. Methods.

3.1.1. Participants.—We tested two groups of infants from the same population as those in Experiment 1, with the younger group being slightly older due to the availability of infants: 15-month-olds ($N = 24$, $M_{age} = 15.0$ months, range = 14.0–16.0, 12 boys) and 17-month-olds ($N = 24$, $M_{age} = 17.62$ months, range = 17.0–18.5, 15 boys). We selected the sample size of 24 before testing began, as in Experiment 1. An additional two 15-month-olds and five 17-month-olds were tested but excluded from the final sample for failure to search on warm-up trials (4), completing fewer than three out of the first four test trials with usable data (2), or experimenter error (1).

3.1.2. Materials & procedure—The stimuli and procedure were identical to those of Experiment 1, with three changes. First, to maximize the infants' attention and engagement in the task, we asked their caregivers to choose one of three small toys (ball, rubber duck, stuffed dog) their child would be most interested in finding. The chosen toy was then used throughout all the trials. Also, we increased the number of test trials to 8, to increase power for finding an effect. Some children became fussy or unwilling to stay on the caregiver's lap as the experiment progressed; for these children, we stopped the experiment after four or six test trials. Most importantly, as described below, we drew the child's attention to both buckets before inviting them to find the toy.

3.1.2.1. Warm-up trials.: The warm-up trials were identical to those of Experiment 1

3.1.2.2. Test trials.: In each of the 8 test trials, the experimenter demonstrated that one bucket was empty by turning it upside down, shaking it, and showing the child the inside of the bucket, and then placed it back in its original position. She *also* lifted and lowered the other bucket, keeping it upright and not revealing its contents, thereby drawing attention to both buckets. The experimenter said, "Look at this!" during both manipulations. The empty bucket was lifted first on half the trials, and second on the other half. After manipulating both buckets, she asked the child to find the toy. Two orders for the location of the toy were constructed – (left, right, right, left, right, left, left, right) and (right, left, left, right, left, right, right, left) – and each order was used for half the children.

3.2. Results

We terminated the study after four test trials for two 17-month-olds and four 15-month-olds, and after six test trials for three 17-month-olds due to fussiness and unwillingness to play further. We also excluded a single test trial for one 17-month-old and two 15-month-olds due to the caregiver releasing them before both buckets had been manipulated (1) or the child failing to approach either of the buckets (2). We used the remaining trials to calculate a percent correct score for each child.

The results of Experiment 2 are summarized in the left panel of Figure 4. We found that 17-month-olds approached the correct bucket on 69% of trials, which was greater than chance ($V = 162$, $p < .001$). Analyzing the 19 17-month-olds who completed at least 7/8 trials, there was no evidence of a learning effect between the first four test trials (66% correct) and the last four test trials (also 66% correct). Furthermore, 17-month-olds were marginally successful on just the first trial (sign test, 17/24 correct choices, $p = .064$), suggesting that their success on this task does not depend on learning to avoid the empty container across trials.

In contrast, 15-month-olds approached the correct bucket on only 49.5% of trials, which was indistinguishable from chance ($V = 80.5$, $p = .843$). Analyzing the 20 infants who completed at least 7/8 trials, there was no evidence of a learning effect between the first four test trials

(51% correct) and the last four test trials (56% correct; V = 31.5, $p = .331$). Comparing the two age groups, 17-month-olds approached the correct bucket significantly more often than 15-month-olds (W = 425, $p = .004$).

To facilitate comparison with Experiment 1, in which we used only 4 test trials, we repeated the above analyses on the first 4 test trials alone. The pattern of results from the first 4 trials fully matched that from all 8 trials: failure at 15 months, success at 17 months, and a significant difference between these age groups. Seventeen-month-olds chose the bucket that contained the toy on 71% of trials, which was better than chance (V = 105, p < .001). This was also significantly better than 17-month-olds' performance in Experiment 1 (W = 396, p = .02). In contrast, 15-month-olds chose the correct bucket on only 48% of trials, which was no different from chance (V = 74, p = .616), and no different than 14-month-olds' performance in Experiment 1 (W = 305.5, p = .719). Seventeen-month-olds again did significantly better than 15-month-olds' (W = 424, p = .004).

To interpret the at-chance performance of the 15-month-olds, we considered the possibility that they were perseverating in their responses across trials. We did find evidence of perseveration: four 15-month-olds chose the same side on every trial, which was greater than chance (binomial test based on 1/128 chance rate: $p < .001$); however, seven 17-month-olds did the same ($p < .001$). When these 12 children were removed from the analysis, the difference between 15- and 17-month-olds' performance was even more apparent: the remaining 15-month-olds chose the correct bucket on 49% of trials, while the remaining 17-month-olds were at 78%, which was significantly better than chance, V = 146, $p = .001$, as well as better than the 15-month-olds, V = 71.5, $p < .001$). Although children were perseverating at both ages, perseveration does not explain the differential pattern of success and failure by age.

A second potential account of the 15-month-olds' performance is that their attention was locked on whichever bucket was manipulated last. Since this was the correct bucket on half the trials and the empty bucket on the other half, that could lead to chance performance. However, 15-month-olds approached the last-manipulated bucket on only 52% of trials, while 17-month-olds approached it on 44% of trials, showing no tendency to approach whichever bucket was manipulated last at either age.

In sum, across both Experiments 1 and 2, we found a consistent failure among 14- and 15-month-old infants to exclude a location on the basis of negation in the 2-cup task, whether the experimenter touched one or both containers. In contrast, 17-month-olds' performance was above chance when their attention was drawn to both buckets in Experiment 2. While there was no effect of the order the buckets were manipulated, it proved important to call 17-month-olds' attention to both buckets rather than only to the empty one. At 23 months (and even at 20 months; see Feiman, et al., 2017), infants succeeded even when only the empty bucket was manipulated, suggesting that the EF capacity necessary to disengage from the sole manipulated bucket in these circumstances is developing over these ages. These ages converge with findings that infants under 17 months fail on those versions of Piaget's invisible displacement task that require disambiguating an object's location (see Gopnik, 1984).

In Experiment 2, we found no evidence that 15-month-olds' poor performance was due to a high rate of perseveration or a tendency to approach the last-manipulated bucket, leaving open the possibility that it was due to a real inability to exclude the empty bucket from consideration. However, alternative explanations for their failure remain open: the youngest infants may have been distracted by the experimenter's manipulation of the buckets, forgotten that a toy was hidden at all, lost interest in finding the toy, or not understood the instruction to find the toy. Experiment 3 addresses these possibilities.

4. Experiment 3. Control task with no exclusion

4.1. Methods

We asked whether 15-month-olds were both able and motivated to find a hidden toy under conditions that were very similar to Experiment 2, but did not require exclusion. In Experiment 3, the hiding procedure allowed children to predict where the toy would be.

4.1.1. Participants—The participants were 24 15-month-olds (M_{age} = 15.0 months, range = 14.1–16.0, 9 boys). We kept the same sample size as in Experiments 1 and 2, which was selected before testing began. An additional 9 infants participated but were excluded from the final sample for failure to search on warm-up trials (8) or parental interference (1).

4.1.2. Procedure—The experimenter held the toy above the screen and caught the child's attention. She then separated her hands *above* the screen, with the toy visible in one hand, and lowered her hands; infants could see which bucket the toy was going towards as it disappeared behind the screen. She then removed the screen, manipulated both buckets as in Experiment 2, and asked the child to find the toy. Because almost 20% of children in Experiment 2 had become fussy before completing 8 test trials, and because children did not behave differently on the first 4 trials alone than on all 8 trials together, we shortened the number of trials back to 4, as in Experiment 1.

The empty bucket was manipulated first on half the trials and last on the other half. The location of the hidden toy varied across trials – either (left, right, right, left) or (right, left, left, right) – and each order was used for half the children.

4.2. Results

For six of the 24 infants, one test trial was excluded due to experimenter error (1) or the child failing to approach either of the buckets (5). Using the remaining test trials, a percent correct score was computed for each child.

The results are depicted in Figure 4. When they could predict where the toy was hidden, 15-month-olds approached the correct bucket on 72% of trials, which was greater than chance, V = 145.5, $p < .001$. They were successful on the first test trial: 19 out of 24 infants approached the correct bucket, sign test, $p = .007$. There was no evidence of a learning effect between the first two test trials (71%) and the last two test trials (73%), V = 20, $p = .790$. Moreover, we find overall success despite evidence that some infants perseverated: 7 out of the 24 infants approached the bucket on the same side across all four test trials, which was greater than chance (binomial test based on 1/8 chance rate: $p = .024$).

Giving infants direct information about the location of the hidden toy had an impact on their searching behavior: the 15-month-olds who participated in Experiment 3 performed better than those in Experiment 2 (V = 426.5, $p = .004$), indicating that 15-month-olds' failure to make an exclusion inference in Experiment 2 was not due to being distracted by the bucket manipulations, nor to their limited memory or motivation. Instead, it was likely due to an inability to use the information that one bucket was empty to exclude it from consideration and direct their search elsewhere.

4.3. Discussion

Experiment 3 rules out multiple alternative explanations for 15-month-olds' failure in Experiment 2. Compared to a looking time study like Cesana-Arlotti and colleagues' (2018; 2020) tasks, the 2-cup task does impose additional demands. It requires the child to decide on and execute an action, which requires maintaining a representation of the ball's location in one bucket as the experimenter lifts the other, choosing between the two buckets, and holding their choice in mind as they leave their caregiver to walk over to the experimenter. Still, these demands were all present in Experiment 3, and 15-month-olds succeeded when they did not need to exclude an option to find the toy. This suggests infants begin to be capable of an exclusion inference in this paradigm between 15 and 17 months and favors the leaner 1-to-1 mapping account of 12-month-olds' success in Cesana-Arlotti and colleagues' task.

In Experiments 4 and 5, we explore infants' capacity for exclusion inferences on the basis of negation in the context of causal reasoning, using the blicket detector paradigm.

5. Experiment 4 – Causal Disambiguation Task: 19-month-olds

Sobel and Kirkham (2006) found that 24-month-olds succeed on indirect screening-off trials (henceforth, "exclusion trials") in a blicket detector paradigm: having seen that A and B together activated the machine, and that A alone did not, they tried the B block alone more often than chance. In contrast, 19-month-olds performed at chance, trying A alone frequently (indeed, as often as they tried B alone). That is, the 19-month-olds showed the signature of failing to exclude an option -- seeing that A was inert did not preclude trying to make A activate the machine as often as trying B. However, their design used only one exclusion trial intermixed among other trial types, and thus could have underestimated infants' abilities. Accordingly, our first experiment in this series is a conceptual replication of Sobel and Kirkham's study with 19-month-olds, expanding the number of exclusion trials from 1 to 4.

Like the 2-cup task, but unlike Cesana-Arlotti, et al.'s (2018; 2020) object individuation tasks, exclusion trials in the blicket detector paradigm cannot be solved by establishing a 1-to-1 mapping between the relevant aspects of the final and the initial displays in a trial. Because the first thing they see are two blocks activating the machine together, infants never start with an unambiguous model of the causal relations between either individual block and the detector. Neither do they end with an unambiguous conclusion -- at the end of an exclusion trial, it remains possible either that block B alone activates the machine or that both blocks A and B must be placed on the machine together in order to activate it. Like the 2-cup task, success on the exclusion trials requires excluding the possibility that the inert block (A) alone activates the machine. Correspondingly, this task allows probing for the signature of failing to make an exclusion inference – putting A alone on the machine, having just seen that A alone is inert.

5.1. Methods

5.1.1. Participants—The participants were 28 19-month-old infants ($N = 28$, M_{age} = 19.79 months, range = 18.71–20.58, 14 boys). We had intended to test 24 children, but accidentally continued testing, and include all the data. An additional 11 infants were excluded for crying (1), failing to act on the first three trials, which led us to abort testing (2), or completing fewer than seven out of ten trials with usable data (4 children did not finish the task; another 4 children used both blocks a majority of the time – a response that we excluded from analysis as uninformative, because it is consistent not only with excluding an option and testing the alternative hypothesis that both blocks are necessary, but also with merely repeating the experimenter's demonstration of successfully activating the machine with both blocks).

5.1.2. Materials and Procedure—The blicket detector was constructed out of a cardboard box and a "magic wand" toy with multicolored LEDs that lit up and spun around when it was activated. The toy could be secretly activated by the experimenter's hand inside the box. We also used a set of wooden blocks; each trial used two new blocks that were identical in shape but differed in color; the shapes of the blocks differed across trials. Thus, in each trial, the child's task was to discover which of the two new blocks activated the machine. The blicket detector and blocks were placed on a tray so the experimenter could slide them in and out of children's reach.

Infants were held on caregivers' laps, across the table from the experimenter. Caregivers were asked to remain silent during the trials. On each trial, the experimenter placed two new blocks (four blocks on the warm-up trial) in front of the blicket detector. She then demonstrated the effect of placing them onto the detector, with the pattern of evidence varying according to the trial type. When demonstrating active blocks, she secretly activated the toy and said "Wow, this makes it go!" When demonstrating inert blocks, she did not activate the toy and said "No, this doesn't make it go." She returned the blocks to their original positions, then slid the blicket detector and blocks towards the child, saying, "Your turn! Can you make it go?"

We coded children's responses by which block they touched to the blicket detector first. Children received feedback on their responses: the toy activated if the child chose an active block or placed both blocks on the detector simultaneously, and did not activate if they chose an inert block. To maintain children's interest, if they chose an inert block on three successive trials, they were encouraged to try the other (active) block at the end of the third trial, but only their first response was included in the data. Each child participated in 10 trials in the following order: one warm-up trial, one no-inference trial, two association trials, two screening-off trials, and four exclusion trials. Figure 6 shows a schematic of no-inference, exclusion, and association trials.

5.1.2.1. Warm-up trial.: Children were introduced to the blicket detector with four blocks in front of it. The experimenter demonstrated blocks in succession: the first block, which activated the toy; then the second block, which did not; then the first and second blocks together twice, which activated the toy; then the third block, which did not; and finally a fourth block, which again activated the toy. The experimenter then removed the first three blocks from the tray, leaving only the fourth (active) block, and asked the child to make it go. On this trial only, caregivers were told that they could encourage their child, and demonstrate the block themselves if their child did not respond immediately.

5.1.2.2. No-inference trial.: The no-inference trial gave infants direct information that one block was active and another was inert. This trial was designed to assess whether infants understood the demonstrations, would remember and use the causal information they had been shown, and could resist simply selecting whichever block the experimenter had manipulated last. The experimenter demonstrated that the first block activated the toy, then that the second block did not, then demonstrated each block again, and finally asked the child to make it go.

5.1.2.3. Association trials.: We included these exploratory trials as a first pass at testing whether children were choosing blocks based on a simple associative strategy: summing up the strength of positive and negative association between each block at the detector's activation. On the first association trial, the experimenter activated the detector with the first block twice, and then with the second block once; the order was reversed for the second association trial. If children were simply adding up associative strength, they should select the block that activated the detector twice. However, if children were reasoning causally, they should conclude that both blocks are active, and would have no particular reason to choose one over the other.

5.1.2.4. Screening-off trials.: These trials scaffolded exclusion inferences by allowing but not requiring it. Children were shown that one of the two blocks was active, and one was not. The experimenter demonstrated that the first block activated the toy, then that the second block did not. She then placed both blocks on the toy simultaneously, which activated it; she did this twice before asking the child to make it go.

5.1.2.5. Exclusion trials.: These trials required that infants exclude the block shown to be inert and choose the block never before demonstrated to activate the machine by itself. The experimenter placed both blocks on the blicket detector simultaneously, which activated it; this ambiguous information was shown twice. Then she placed one of the blocks on the detector by itself, which did not activate it, before asking the child to make it go. If infants excluded the inert block from consideration, they should either try both blocks together or the other block alone. In these trials, the child was never shown that one of the blocks alone activated the toy.

5.2. Results

We excluded seven trials across five infants because the child failed to make a response (4) or touched both blocks to the blicket detector simultaneously (3 trials: 2 exclusion, 1 screening-off trial). Using the remaining trials, the percentage of correct choices was computed for each trial type. Results are summarized in Figure 5.

5.2.1. No-inference trial.—On the single no-inference trial, 21 out of 28 infants tried to activate the detector using the active block, which was greater than chance on a sign test, $p = .013$. This demonstrates that they were motivated to activate the blicket detector, could use the causal information we showed them to make their choice, and could inhibit using the last block that the experimenter used.

5.2.2. Association trials.—On association trials, infants chose the block that had activated the detector twice on 55.8% of trials, which was no different than chance ($V = 20$, $p = .299$). This gives no sign that infants were relying on an associative strategy. Furthermore, the associative strategy predicts success on the screening-off trials, where infants failed.

5.2.3. Screening-off trials.—Surprisingly, on screening-off trials – which provided infants with direct information about the causal status of each block – infants performed at chance: they selected the active block on 50% of trials, $V = 10.5$, $p = 1$. It is possible that the ambiguous demonstration of both blocks on the blicket detector together, *after* they had seen the demonstrations of the active and inactive blocks, caused them to forget the information they had seen earlier. At any rate, the failure at screening-off trials at 19-months of age is not robust; Sobel and Kirkham's 19-month-olds succeeded at this trial type. Further research would be needed to explore what factors led our children to fail, and theirs to succeed. Because our interest is in the earliest age of success on exclusion inferences, we dropped the screening-off trials from Experiment 5.

5.2.4. Exclusion trials.—Finally, unlike Sobel and Kirkham (2006), we found that 19-month-olds succeeded at exclusion trials. This is likely because we used a block of four exclusion trials, rather than a single exclusion trial intermixed among other types of trials. Infants chose the active block on 65% of trials, which was greater than chance, $V = 148.5$, $p = .005$. There was no evidence of a learning effect between the first two exclusion trials (66.1%) and the last two exclusion trials (63.0%) $V = 42$, $p = .83$. Furthermore, 21 out of 28 infants selected the active block on their first exclusion trial, sign test, $p = .013$, indicating that infants spontaneously exclude a candidate cause of an event by 19 months of age.

Having found success at 19-months of age on exclusion trials, unlike Sobel and Kirkham (2006), we now look for the youngest age of success. Based on the results of Experiments 2 and 3, we target 15- and 17-month-olds. We first must confirm that children of both ages succeed on the no-inference trials, establishing that they understand the blicket detector task. Then, if the difference between 15- and 17-month-olds in Experiment 2 reflects an emerging capacity for exclusion, 17-month-olds should succeed at the exclusion trials, while 15-month-olds should fail, choosing the block just shown to be inert as often as the block they had never seen activating the machine on its own. Of course, the blicket detector task may be harder than the 2-cup task overall, because the evidence presented is consistent both with B alone having the causal power to activate the machine, or with both A and B together being necessary. Still, if Sobel and Kirkham's data underestimated 19-month-olds' capacity for exclusion inferences on exclusion trials, and if 15-month-olds lack this capacity, we should find a significant effect of age as infants' performance improves from failure at 15 months to success at 19 months.

6. Experiment 5: Causal task with 15- and 17-month-olds

6.1. Methods

6.1.1. Participants—With 24 19-month-old infants, Experiment 4 found a medium-sized effect on the exclusion trials (Cohen's $d = 0.60$). A power analysis revealed that 90% power for detecting an effect of this size in a one-sample t-test would require 30 participants. Since the blicket detector method is not commonly used with infants this young, prior to testing we decided to increase our sample size to 36 infants in each age group, which comfortably exceeds the required sample of 30. We tested two groups of infants: 15-month-olds ($N = 36$, $M_{age} = 14.88$ months, range = 14.08–15.59, 22 boys) and 17-month-olds ($N = 36$, $M_{age} = 17.83$ months, range = 17.34–18.45, 21 boys). An additional twelve 15-month-olds and four 17-month-olds were tested but excluded from the final sample for parental interference (1), for completing fewer than seven out of ten trials with usable data (8; 4 children did not finish the task and another 4 used both blocks most of the time), or for making no response on the first three trials, which led us to abort testing (7).

6.1.2. Materials & procedure—Each child participated in 10 trials: one warm-up trial, three no-inference trials, and four exclusion trials, as well as two association trials, always presented in this order. Screening-off trials were removed, while Association and Exclusion trials remained unchanged. The procedure for each trial was identical to Experiment 4, except for changes detailed below, which were intended to shorten the experiment for younger participants (Figure 6).

6.1.2.1. Warm-up trial.: In an effort to decrease the likelihood of putting two blocks on the machine together, the experimenter did not demonstrate the first and second blocks on the toy together during the warm-up trial.

6.1.2.2. No-inference trials.: On each trial, the experimenter performed the demonstration sequence once (first block activates machine, second does not) and the number of trials was increased from 1 to 3 to obtain a more robust assessment of infants' understanding that some blocks activate the toy and some do not, of their understanding that they were to try to activate the machine, and their ability to inhibit imitating the last action of the experimenter.

6.2. Results

We excluded 12 trials across five 15-month-olds and five 17-month-olds due to the child's failure to make a response (10) or parental interference (2). We also excluded 18 trials across 11 15-month-olds (4 no-inference; 6 exclusion; 8 association) and 4 trials across two

17-month-olds (1 no-inference; 2 exclusion; 1 association) due to the child touching both blocks to the blicket detector simultaneously. The results are summarized in Figure 7.

6.2.1. No-inference trials.—

Both groups of participants succeeded on the no-inference trials. Seventeen-month-olds tried to activate the blicket detector using the active block on 62.6% of trials, which was greater than chance ($V = 540$, $p < .001$). Similarly, 15-month-olds chose the active block on 59.3% of trials, which was greater than chance ($V = 453.5$, $p < .001$), and not significantly different from the 17-month-olds ($W = 582.5$, $p = .43$). Infants at both ages were motivated to activate the blicket detector, made choices based on the causal information they witnessed, and could switch their attention away from the inert block that was manipulated last.

6.2.2. Exclusion trials.—

Seventeen-month-olds also succeeded at exclusion trials. They chose the active block on 60% of trials, which was greater than chance ($V = 191.5$, $p = .03$). There was no evidence of a learning effect between the first two exclusion trials (58%) and the last two exclusion trials (63%, $V = 56$, $p = .51$; one infant generated no usable data on the last two exclusion trials and was dropped from this comparison). Unlike the 17-month-olds, 15-month-olds failed on the exclusion trials, choosing the active block on only 52.8% of the trials, which was indistinguishable from chance ($V = 112.5$, $p = .238$). However, a Mann-Whitney test found that these age groups were not significantly different from each other ($W = 534$, $p = .178$).

To compare responses on exclusion trials across the whole age range, we combined responses from the 15- and 17-month-olds in Experiment 5 with the 19-month-olds from Experiment 4. A Kruskall-Wallis test found a marginally significant effect of age on infants' performance ($\chi^2(2) = 5.14$, $p = 0.076$). A post-hoc logistic regression treating age as a continuous variable found a significant effect of age, with older children more likely to choose the active block ($\chi^2(2) = 5.98$, $p = 0.014$). Similarly, planned Mann-Whitney tests comparing the three age groups found that the 17-month-olds were not significantly different from the 19-month-olds ($W = 440.5$, $p = 0.374$), while the 15- and 19-month-olds differed significantly from each other ($W = 344.5$, $p = 0.023$).

A final analysis explored whether perseveration might explain the failures of the younger children. Indeed, we found that 14 15-month-olds selected the same-side block across the nine trials, which was much greater than chance (binomial test based on 1/256 chance rate: $p < .001$); only four 17-month-olds and four 19-month-olds in Experiment 4 did the same, which was also greater than chance ($p < .001$). However, when we removed these children from the analysis, the results were essentially unchanged. The remaining 24 19-month-olds chose the active block on 79% of no-inference trials ($V = 237.5$, $p = 0.004$) and 68% of exclusion trials ($V = 148.5$, $p = 0.005$) and the remaining 32 17-month-olds chose the active block on 62.5% of no-inference trials, ($V = 444$, $p < 0.001$), and 61.5% of exclusion trials, ($V = 191.5$, $p = .03$). In contrast, the remaining 22 15-month-olds chose the active block on 63% of no-inference trials, ($V = 202$, $p = .002$), but only 54.5% of exclusion trials, ($V = 89.5$, $p = .26$). These remaining 15- and 17-month-olds were still not significantly different from each other, while the 15-month-olds performed marginally worse than the 19-month-olds in Experiment 4 ($W = 185$, $p = 0.074$).

6.2.3. Association trials.—

Finally, we found no evidence of any consistent preference on the exploratory association trials: 17-month-olds chose the block that activated the blicket detector twice on 47.1% of trials (V = 32.5, $p = .594$), and 15-month-olds chose it on 58.6% of trials, (V = 75, $p = .117$; one 15-month-old and one 17-month-old generated no usable data on association trials and were dropped from these comparisons). There was no indication that infants were using a simple strategy of adding up associative strength when choosing which block to try. However, given that there were only two association trials, and that these trials came at the end of a relatively long study, these null findings may not be meaningful. The use of a simple associative strategy has been ruled out for similar trials in 4-year-olds (Sobel, Tenenbaum, & Gopnik, 2004), but it is still possible that it plays a role in infants' performance.

6.3. Discussion

The blicket detector task is clearly difficult at this age – to our knowledge, no published paper employs this design with infants younger than 18 months. As noted above, Sobel and Kirkham (2006) found that 19-month-olds performed at chance on a single exclusion trial. The present experiments establish that 19-month-olds can successfully exclude a block from consideration in a causal context, forming an action plan on the basis of that inference. However, we found no evidence that 15-month-olds can do so, and while the 17-month-olds performed above chance as a group, their performance did not differ significantly from that of the 15-month-olds.

Still, the difficulty of the task alone cannot explain 15-month-olds' failure. Based on their success on no-inference trials, 15-month-olds' failure to exclude the inert block on the exclusion trials cannot be due to a lack of motivation or difficulty understanding the causal demonstrations. Furthermore, on both trial types, the experimenter's final demonstration was of the inert block; since infants successfully switched away from this block on no-inference trials, it is unlikely that difficulty with attention-switching could explain 15-month-olds' chance performance on exclusion trials. Even after excluding those who perseverated by always choosing a block on the same side, the remaining 15-month-olds did no better than chance. We conclude that we have narrowed the age of earliest success on exclusion inferences in the blicket detector paradigm to between 17 and 19 months of age.

7. General Discussion

Our results extend previous research on the exclusion component of the disjunctive syllogism in three ways. First, from non-human animals to human infants in Call's (2004) 2-cup task; second, from the 2-cup task to the blicket detector task; and third, by establishing the human developmental trajectory in both paradigms. Experiments 1, 2, and 3 established the developmental course of success on the 2-cup task. A clear developmental pattern emerged: failure at 14 and 15 months (Experiments 1 and 2); fragile success at 17 months (failure in Experiment 1; success in Experiment 2); with increasingly robust success up through 20 and 23 months (Feiman et al., 2017; Mody and Carey, 2016). In Experiments 4 and 5, we observed the same developmental trajectory in the blicket detector paradigm, investigating trials that require an exclusion inference: failure at 15 months, fragile success at 17 months (success relative to chance, but not significantly better than 15-month-olds' performance), and significantly greater success by 19 months.

Extending findings of successful exclusion inferences from the 2-cup task to the blicket detector task is important for several reasons. First, unlike both Cesana-Arlotti, et al.'s paradigm and the 2-cup task, perception never provides an unambiguous initial model of the causal properties of the blocks and the unfolding evidence does not force the conclusion that one of the blocks is inert (because using both blocks together could be necessary). Thus, children avoiding that block provides strong evidence that they are excluding it from consideration without sufficient evidence to settle on a positive characterization of exactly how the blocks and the detector work; that is, without the possibility of establishing a 1-to-1 mapping of causal properties to individual objects. Finally, the two paradigms differ also in whether the content being reasoned about are objects' locations or invisible causal properties. This means that the convergence in the developmental time course of both tasks is consistent with a content-independent capacity for exclusion emerging at these ages, and content independence is a hallmark of logically structured thought.

7.1. Resolving the conflicting data about representations of negation in infancy.

Fully charting the emergence of negation and disjunction within logically structured thought requires a better understanding of why children solve different tasks that plausibly require exclusion at very different ages, ranging from 1 to 4 years. On the one hand, some mutual exclusivity tasks (Jin & Song, 2017; Pomiechowska, et al., 2021; Yin & Csibra, 2015) and Cesana-Arlotti, et al.'s (2018, 2020) identity disambiguation paradigm find infants succeeding at 12 months of age. On the other hand, many experiments using various location disambiguation paradigms show that children under four years of age do not reach the certain conclusion that the disjunction in reasoning by exclusion produces (Call and Carpenter, 2001; Grigoroglou, et al., 2019; Mody & Carey, 2016, but see Gautam, et al., 2021, and Leahy, under review, for further discussion). Adding to this, we have found that 15-month-olds do not use negation to exclude an option in either the 2-cup or the blicket detector task.

There are two possible resolutions of these conflicting data. One possibility is that younger infants can reason by exclusion (representing both *not* and *or*), but that this capacity is masked by too-high performance demands being imposed by those tasks on which they fail. The other possibility is that younger infants' successes do not depend upon reasoning by exclusion, and that such reasoning emerges in two steps: first, exclusion based on negation beginning around 17–19 months of age, and second, the capacity to relate multiple options through a disjunctive premise beginning around age 4 (see Leahy & Carey, 2020). We focus the remaining discussion on the target of the present experiments: exclusion on the basis of negation.

We take both resolutions of the conflicting data to be open possibilities. Deciding between them requires formulating and testing alternative explanations of younger infants' successes on some tasks, as well as specific hypotheses about how performance demands might explain older infants' failures on others (for an example of this approach in a different case study, see Diamond, 1990; 2006). Both of these are important targets for future research. An important and related question is why 15-month-old children fail to make exclusion inferences in the 2-cup task, while many non-human animals succeed. Studying the developmental course of success in those species, and relating it to the development of their executive capacities, may help to understand how performance demands affect this task (see Diamond, 1990). Alternatively, if the developmental courses of success in humans and non-human animals differ, then we should explore whether and how the bases of success might differ as well.

The present studies included several tests of performance demands that might be limiting success. Experiment 2 yielded evidence that a specific performance limitation – difficulty disengaging attention from the manipulated bucket – prevents 17-month-olds from succeeding in one version of the 2-cup task. However, Experiment 3 provided further evidence against this and other performance limitations explaining the failure of 15-month-olds, including their having insufficient working memory to remember that the toy is behind the occluder, insufficient motivation to find the toy, a tendency to perseverate, or an inability to inhibit attention to the last manipulated item. Still, the possibility that other performance demands are responsible for the failure of 15-month-olds in the 2-cup and blicket detector paradigms remains. A target for future research is formulating and testing hypotheses about what these might be.

7.2. Alternative hypotheses for successes on tasks that might involve negation.

It is, however, also possible that younger infants' successes on other tasks (Cesana-Arlotti, et al., 2018; 2020, Jin & Song, 2017; Pomiechowska, et al., 2021; Yin & Csibra, 2015) have alternative explanations that do not draw on negation. In this case, older infants' failures may really reflect an absence of negation in logically structured thought. Alternative explanations of early success can fall into two general categories. First, some successes may rely on positive representations alone. In the introduction, we discussed these kinds of alternative explanations for success in referent disambiguation tasks (i.e. mutual exclusivity phenomena in word learning) and in Cesana-Arlotti, et al.'s identity disambiguation tasks. The 1-to-1 mapping processes we described there did not rely on any form of negation. Second, some successes may also rely on *implicit* representations of negation, which differ from negation in logically structured thought in principled ways.

7.2.1. Implicit representations of negation.—Of course, positing that a representation of negation is implicit only counts as an alternative if it is possible to specify how an implicit representation differs from an explicit one. One way to characterize implicit representations of negation is as internal to, and encapsulated within, particular computations; i.e. as *domain specific*. Such implicit representations would be tokened independently at each implementation, lack a common content-neutral reusable symbol (a logical operator *not*), and may not be available for deployment more broadly in flexible logically structured thought (for similar suggestions, see Bermudez, 2003; Bohn, et al., 2020; Völter & Call, 2017). A second way in which representations of negation might be implicit is if they are capable of fulfilling only part of the function of the logical operator, but cannot do everything that logical *not* does; i.e. if they have only *partial function*.

There is significant independent evidence that computations of negation that are implicit in both of these senses are ubiquitous in both human and non-human cognition (e.g. Kuffler, 1953; Rao & Ballard, 1999), and they could plausibly underlie success on some of the tasks we have discussed. For example, in violation of expectancy paradigms, *error signals* that indicate a mismatch between what an infant expects to see and what they actually see are an implicit form of negation in both senses: they could be specific to object recognition or visual processing (domain specificity), and they carry no commitment that the expected and actual outcomes are incompatible, which fails to fully implement the contrariety function of negation (partial function). Similarly, in the referent and identity disambiguation tasks, the constraints that play a role in 1-to-1 mapping may *inhibit* one another. Activating one representation could reduce the activation, or the probability of activation, of another without being completely incompatible with it (the same partial function again). In referent disambiguation, knowing a label for an object might inhibit mapping a novel label to that object. In identity disambiguation, seeing a ball emerge from behind a screen might inhibit representing that the same ball is in a different location.

In some cases, it may be possible to find positive evidence that a representation of negation is implicit. Hochmann and Toro (2021) recently reported evidence that 11-month-old infants can learn to discriminate strings of up to five syllables that are all the same from strings in which the last syllable is different from (perhaps represented as *not the same as*) the preceding ones. However, in follow-up work, Hochmann (under review) finds positive evidence that the form of negation in this task is very limited -- infants fail to discriminate longer strings following the same pattern, suggesting that the negation can only be deployed to relate two syllables that are both held in auditory working memory, and do not represent strings where one syllable is different as *not all same.*

The key point is that evidence for implicit negation does not entail evidence for the kind of negation that exists in logically structured thought, the kind that supports flexible and content-independent reasoning. There is no reason to assume that every animal capable of one kind of inhibition (e.g. having a center-surround inhibition architecture in the retina; Kuffler, 1953) should thereby be capable of representing the kind of negation that is the meaning of words like “no” and “not” (see Horn, 1989), and which is involved in composing complex, effable thoughts. This means that even unambiguous evidence that a function of negation is deployed in one task cannot by itself establish a capacity for logically structured thought more broadly. Success on any one task could always involve negation that is implicit in one or both senses.

Of course, the same point applies to each of the studies we reported here: it is possible that infants’ successes on both location and causal disambiguation tasks rely on implicit forms of negation. To pinpoint the emergence of negation of that content-independent variety that is required for logically structured thought, one strategy is to find evidence of a relation between uses of negation in different tasks, which otherwise differ in as many ways as possible. Here, we found convergence in the course of development across location disambiguation tasks, both Call’s 2-cups tasks and Piagetian invisible displacement tasks on the one hand, and exclusion trials within the blicket detector paradigm on the other.

Nevertheless, finding similar ages of success is only circumstantial evidence for logically structured thought. Co-incidences of ages of success across tasks could be just coincidences. Stronger evidence that they reflect the emergence of a common capacity can come from within-individual correlations between tasks. One example of such evidence comes from findings that toddlers' production of "gone" or "all gone" was specifically correlated with performance on the variety of Piagetian invisible displacement tasks that involves exclusion, as opposed to other Piagetian milestones that were specifically related to other verbal expressions (Corrigan, 1978; Gopnik, 1984; Gopnik & Meltzoff, 1984, 1985, 1986; McCune-Nicolich, 1981; Tomasello & Farrar, 1984). The convergence between when children begin to produce these words and their successes on invisible displacement tasks, on the 2-cup task, and on exclusion trials of the blicket detector task is consistent with the hypothesis that a common developmental change having to do with the availability of explicit symbols for negation in logically structured thought may be underway around this age.

Future research may look for additional evidence of within-individual correlations between other measures of negation. For instance, are children who succeed on one task more likely to succeed on the others, controlling for shared performance factors like attention and inhibitory control? Even better evidence would be if success or training on one task causally affects performance on another -- a kind of structural priming approach (Pickering & Ferreira, 2008) applied to the study of infants. This approach could reveal that some tasks affect each other, but not others. Future research may be able to chart which tasks recruit a common symbolic representation of negation, and which recruit either more implicit forms, or perhaps are solved on the basis of positive representations alone.

Finally, the distinction between implicit and explicit negation raises the question of whether and how the two might be related, and whether implicit forms may serve as a developmental precursor to the explicit logical symbol. Indeed, the title of Cesana-Arlotti, et al. (2018) is *Precursors of logical reasoning in preverbal human infants*. Though they do not specify what distinguishes a precursor from the logical symbol, the characterization of implicit negation above offers a set of testable hypotheses for future research. If, for example, a form of implicit negation that is specialized for object tracking underlies infants' success on Cesana-Arlotti, et al.'s tasks, it could be a developmental precursor that plays a causal role in the acquisition of explicit, symbolic negation in logically structured thought later in development.

8. Conclusion

Across five experiments and two paradigms, we have explored the ontogenesis of the role of negation in reasoning by exclusion -- the capacity to exclude one option from consideration. We found infants succeeding between 17 and 19 months of age, but failing at 15 months, even as controls showed that many non-logical aspects of both tasks were within their grasp. These findings conflict with evidence of younger success on different tasks. The resolution of these conflicting data await further research aimed at adjudicating between alternative explanations of both the successes and the failures. Still, such conflicting results raise an intriguing possibility. Different, sometimes partial, instantiations of logical functions may exist in distinct, content-specific forms independently from each other and may each be recruited by different tasks. These may precede the emergence of domain-neutral, content-independent logical operators of the kind involved in adults' logically structured thought.

Acknowledgements

We thank Allison Hyland, Evelyn Xing, Haven Jones, Luke Anderson, Sara Al Mughairy, Courtney Nugent, and Stella Gerdemann for help with data collection. We also thank Renee Baillargeon in her role of editor, Ned Block, Kate Davidson, Rhea Howard, Masoud Jasbi, Brian Leahy and two anonymous reviewers for helpful comments and discussion, as well as Heather Yu for assistance with preparing the final manuscript. This research was supported by NIH grant #RO1-HD038338 and McDonnell Foundation grant #220020418.

References

Bermúdez J (2003). Thinking without Words. Oxford University Press. DOI:10.1093/acprof:oso/9780195159691.001.0001

Bion RAH, Borovsky A, & Fernald A (2012). Fast mapping, slow learning: Disambiguation of novel word-object mappings in relation to vocabulary learning at 18, 24, and 30 months. Cognition, 126, 39–53. [PubMed: 23063233]

Bohn M, Call J, & Völter CJ (2020). Evolutionary Precursors of Negation in Non-Human Reasoning. In The Oxford Handbook of Negation.

Burge T (2010). Steps toward origins of propositional thought. Disputatio, 4(29), 39–67.

Byers-Heinlein K, & Werker JF (2009). Monolingual, bilingual, trilingual: Infants' language experience influences the development of a word-learning heuristic. Developmental Science, 12(5), 815–823. [PubMed: 19702772]

Call J (2004). Inferences about the location of food in the great apes (*Pan paniscus, Pan troglodytes, Gorilla gorilla, and Pongo pygmaeus*). Journal of Comparative Psychology, 118, 117–128.

Call J, & Carpenter M (2001). Do apes and children know what they have seen? Animal Cognition, 3(4), 207–220.

Carey S (2009). Where our number concepts come from. The Journal of Philosophy, 106(4), 220. [PubMed: 23136450]

Carruthers P (2002). The cognitive functions of language. Behavioral and brain sciences, 25(6), 657–674. [PubMed: 14598623]

Cesana-Arlotti N, Kovács ÁM, & Téglás E (2020). Infants recruit logic to learn about the social world. Nature Communications, 11(1), 5999.

Cesana-Arlotti N, Martín A, Téglás E, Vorobyova L, Cetnarski R, & Bonatti LL (2018). Precursors of logical reasoning in preverbal human infants. Science, 359(6381), 1263–1266. [PubMed: 29590076]

Clark A (2006). Language, embodiment, and the cognitive niche. Trends in cognitive sciences, 10(8), 370–374. [PubMed: 16843701]

Corrigan R (1978). Language development as related to stage 6 object permanence development. Journal of Child Language, 5(2), 173–189.

Davidson D (1999). The emergence of thought. Erkenntnis, 51(1), 511–521.

Descartes R (1649). Les passions de l'ame, Amsterdam, Lodewijk Elsevier, and Paris, Henry le Gras. In: Adam & Tannery (1964–74), vol. XI.

Diamond A (1990). Developmental time course in human infants and infant monkeys, and the neural bases of inhibitory control in reaching. Annals of the New York Academy of Sciences, 608, 637–676. [PubMed: 2075965]

Diamond A, Prevor MB, Callender G, & Druin DP (1997). Prefrontal cortex cognitive deficits in children treated early and continuously for PKU. Monographs of the society for research in child development, i–206.

Diamond A (2006). The Early Development of Executive Functions. In Bialystok E & Craik FIM, Eds. Lifespan Cognition: Mechanisms of Change, (pp. 286–308). Oxford, England: Oxford University Press.

Diamond A (2013). Executive functions. Annual review of psychology, 64, 135. doi:10.1146/annurev-psych-113011-143750.

Ekramnia M, Mehler J, & Dehaene-Lambertz G (2021). Disjunctive inference in preverbal infants. iScience, 24(10), 103203. [PubMed: 34703998]

Feigenson L & Carey S (2005). On the limits of infants' quantification of small object arrays. Cognition, 97, 295–315. [PubMed: 16260263]

Feigenson L, Carey S, & Hauser M (2002). The representations underlying infants' choice of more: object files vs. analog magnitudes. Psychological Science, 13, 150–156. [PubMed: 11933999]

Feiman R, Mody S, Sanborn S, & Carey S (2017). What do you mean, no? Toddlers' comprehension of logical "no" and "not". Language Learning and Development, 13(4), 430–450.

Fodor JA (1975). The language of thought (Vol. 5). Harvard university press.

Gautam S, Suddendorf T, & Redshaw J (2021). When can young children reason about an exclusive disjunction? A follow up to Mody and Carey (2016). Cognition, 207, 104507. [PubMed: 33203586]

Golinkoff RM, Hirsh-Pasek K, Bailey LM, & Wenger NR (1992). Young children and adults use lexical principles to learn new nouns. Developmental Psychology, 28(1), 99.

Grigoroglou M, Chan S, & Ganea PA (2019). Toddlers' understanding and use of verbal negation in inferential reasoning search tasks. Journal of experimental child psychology, 183, 222–241. [PubMed: 30913424]

Gopnik A (1984). The acquisition of gone and the development of the object concept. Journal of Child Language, 11(2), 273–292. [PubMed: 6746777]

Gopnik A, & Meltzoff AN (1984). Semantic and cognitive development in 15-to 21-month-old children. Journal of Child Language, 11(3), 495–513. [PubMed: 6501461]

Gopnik A, & Meltzoff AN (1985). From people, to plans, to objects: Changes in the meaning of early words and their relation to cognitive development. Journal of Pragmatics, 9(4), 495–512.

Gopnik A, & Meltzoff AN (1986). Relations between semantic and cognitive development in the one-word stage: The specificity hypothesis. Child Development, 1040–1053.

Gopnik A, Sobel DM, Schulz LE, & 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. [PubMed: 11552758]

Halberda J (2003). The development of a word-learning strategy. Cognition, 87, B23–B34. [PubMed: 12499109]

Halberda J (2006). Is this a dax which I see before me? Use of the logical argument disjunctive syllogism supports word-learning in children and adults. Cognitive psychology, 53(4), 310–344. [PubMed: 16875685]

Hill A, Collier-Baker E, & Suddendorf T (2012). Inferential reasoning by exclusion in children (*Homo sapiens*). Journal of Comparative Psychology, 126(3), 243–254. [PubMed: 21728410]

Hjelmfelt A, Weinberger ED, & Ross J (1991). Chemical implementation of neural networks and Turing machines. Proceedings of the National Academy of Sciences, 88(24), 10983.

Hochmann J-R (under review). Representations of abstract relations in infancy.

Hochmann JR, & Toro JM (2021). Negative mental representations in infancy. Cognition, 104599. [PubMed: 33526259]

Horn L (1989). A natural history of negation. Chicago, IL: University of Chicago Press.

Horn L (2015). On the contrary: Disjunctive syllogism and pragmatic strengthening. In The road to universal logic (pp. 241–265). Birkhäuser, Cham.

Jasbi M, Bohn M, Long B, Fourtassi A, Barner D, & Frank MC (2019). Comment on Cesana-Arlotti et al. (2018). PsyArXiv.

Jin K, & Song H (2017). You changed your mind! Infants interpret a change in word as signaling a change in an agent's goals. Journal of Experimental Child Psychology, 162, 149–162. [PubMed: 28605696]

Kibbe MM (2015). Varieties of visual working memory representation in infancy and beyond. Current Directions in Psychological Science, 24(6), 433–439.

Kuffler SW (1953). Discharge patterns and functional organization of mammalian retina. Journal of neurophysiology, 16(1), 37–68. [PubMed: 13035466]

Leahy BP (under review). Don't you see the possibilities? Evidence that young preschoolers lack possibility concepts.

Leahy BP, & Carey SE (2020). The acquisition of modal concepts. Trends in Cognitive Sciences, 24(1), 65–78. [PubMed: 31870542]

Leslie AM, Xu F, Tremoulet PD, & Scholl BJ (1998). Indexing and the object concept: developing 'what' and 'where' systems. Trends in cognitive sciences, 2(1), 10–18. [PubMed: 21244957]

Markman EM, & Wachtel GF (1988). Children's use of mutual exclusivity to constrain the meanings of words. Cognitive Psychology, 20(2), 121–157. [PubMed: 3365937]

Mather E, & Plunkett K (2010). Novel labels support 10-month-olds' attention to novel objects. Journal of Experimental Child Psychology, 105(3), 232–242. [PubMed: 20031152]

Mather E, & Plunkett K (2012). The role of novelty in early word learning. Cognitive Science, 36(7), 1157–1177. [PubMed: 22436081]

McCulloch WS, & Pitts W (1943). A logical calculus of the ideas immanent in nervous activity. The Bulletin of Mathematical Biophysics, 5(4), 115–133.

McCune-Nicolich L (1981). The cognitive bases of relational words in the single word period. Journal of Child Language, 8(1), 15–34. [PubMed: 7204519]

Merriman WE, & Bowman LL (1989). The mutual exclusivity bias in children's word learning. Monographs of the Society for Research in Child Development, 54(3–4, serial no. 220).

Mervis CB, & Bertrand J (1994). Acquisition of the novel name–nameless category (N3C) principle. Child Development, 65(6), 1646–1662. [PubMed: 7859547]

Moher M, Feigenson L, & Halberda J (2010). A one-to-one bias and fast mapping support preschoolers' learning about faces and voices. Cognitive Science, 34(5), 719–751. [PubMed: 21564234]

Mody S, & Carey S (2016). The emergence of reasoning by the disjunctive syllogism in early childhood. Cognition, 154, 40–48. [PubMed: 27239748]

Penn DC, Holyoak KJ, & Povinelli DJ (2008). Darwin's mistake: Explaining the discontinuity between human and nonhuman minds. Behavioral and Brain Sciences, 31(02), 109–130. [PubMed: 18479531]

Pepperberg IM, Gray SL, Mody S, Cornero FM, & Carey S (2018). Logical reasoning by a Grey parrot? A case study of the disjunctive syllogism. Behaviour, 156(5–8), 409–445.

Pepperberg IM, Koepke A, Livingston P, Girard M, & Hartsfield LA (2013). Reasoning by inference: Further studies on exclusion in grey parrots (*Psittacus erithacus*). Journal of Comparative Psychology, 127(3), 272–281. [PubMed: 23421751]

Pickering MJ, & Ferreira VS (2008). Structural priming: a critical review. Psychological bulletin, 134(3), 427–459. [PubMed: 18444704]

Plotnik JM, Shaw RC, Brubaker DL, Tiller LN, & Clayton NS (2014). Thinking with their trunks: Elephants use smell but not sound to locate food and exclude nonrewarding alternatives. Animal Behaviour, 88, 91–98.

Pomiechowska B, Bródy G, Csibra G, & Gliga T (2021). Twelve-month-olds disambiguate new words using mutual-exclusivity inferences. Cognition, 104691. [PubMed: 33934847]

Premack D, & Premack AJ (1994). Levels of causal understanding in chimpanzees and children. Cognition, 50(1–3), 347–362. [PubMed: 8039368]

Premack D (1995). Cause/induced motion: Intention/spontaneous motion. In Changeux JP & Chavaillon J (Eds.), Origins of the human brain (pp. 286–308). Oxford, England: Oxford University Press.

Rao RPN, & Ballard DH (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. Nature Neuroscience, 2(1), 79–87. [PubMed: 10195184]

Sabbatini G, & Visalberghi E (2008). Inferences about the location of food in capuchin monkeys (*Cebus apella*) in two sensory modalities. Journal of Comparative Psychology, 122, 156–166. [PubMed: 18489231]

Schloegl C, Schmidt J, Boeckle M, Weiß BM, & Kotrschal K (2012). Grey parrots use inferential reasoning based on acoustic cues alone. Proceedings of the Royal Society B, 279(1745), 4135–4142. [PubMed: 22874753]

Scholl BJ (2001). Objects and attention: The state of the art. Cognition, 80(1–2), 1–46. [PubMed: 11245838]

Sobel DM, & Kirkham NZ (2006). Blickets and babies: The development of causal reasoning in toddlers and infants. Developmental Psychology, 42(6), 1103–1115. [PubMed: 17087545]

Sobel DM, Tenenbaum JB, & Gopnik A (2004). Children's causal inferences from indirect evidence: Backwards blocking and Bayesian reasoning in preschoolers. Cognitive Science, 28(3), 303–333.

Spelke ES, & Kinzler KD (2007). Core knowledge. Developmental science, 10(1), 89–96. [PubMed: 17181705]

Stavans M, Lin Y, Wu D, & Baillargeon R (2019). Catastrophic individuation failures in infancy: A new model and predictions. Psychological review, 126(2), 196. [PubMed: 30550314]

Strawson PF (1950). On Referring. Mind, 59(235), 320–344.

Tomasello M, & Farrar MJ (1984). Cognitive bases of lexical development: Object permanence and relational words. Journal of Child Language, 11(3), 477–493. [PubMed: 6501460]

Völter CJ, & Call J (2017). Causal and inferential reasoning in animals. APA Handbook of Comparative Psychology: Perception, Learning, and Cognition, Vol. 2, 643–671. 10.1037/0000012-029

Waismeyer A, Meltzoff AN, & Gopnik A (2015). Causal learning from probabilistic events in 24-month-olds: an action measure. Developmental science, 18(1), 175–182. [PubMed: 25041264]

Walker CM, & Gopnik A (2014). Toddlers infer higher-order relational principles in causal learning. Psychological science, 25(1), 161–169. [PubMed: 24270464]

Watson JS, Gergely G, Csanyi V, Topal J, Gacsi M, & Sarkozi Z (2001). Distinguishing logic from association in the solution of an invisible displacement task by children (Homo sapiens) and dogs (Canis familiaris): Using negation of disjunction. Journal of Comparative Psychology, 115(3), 219. [PubMed: 11594490]

White KS, & Morgan JL (2008). Sub-segmental detail in early lexical representations. Journal of Memory and Language, 59(1), 114–132.

Yin J, & Csibra G (2015). Concept-based word learning in human infants. Psychological Science, 26(8), 1316–1324. [PubMed: 26195636]

Yurovsky D, & Frank MC (2017). Beyond naïve cue combination: Salience and social cues in early word learning. Developmental Science, 20(2), e12349.

Zosh JM, & Feigenson L (2009). Beyond 'what' and 'how many': capacity, complexity and resolution of infants' object representations. The origins of object knowledge, 25–51.

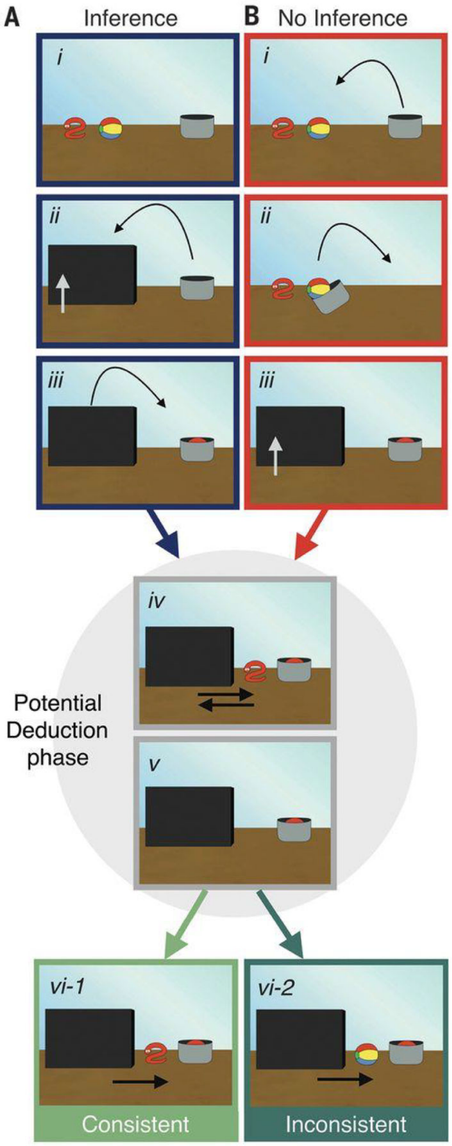

Figure 1.

Schematic of the procedure used by Cesana-Arlotti, et al. (2018). Condition (Inference vs. No-Inference) was manipulated between subjects, while test trial (Consistent vs. Inconsistent) was manipulated within subject. In the No-Inference condition, infants could see which object the cup scooped, and so knew the locations of both the snake and the ball throughout. In the Inference condition, the screen rose and hid the object scooped by the cup prior to the potential deduction phase. Reprinted from Cesana-Arlotti, et al. (2018).

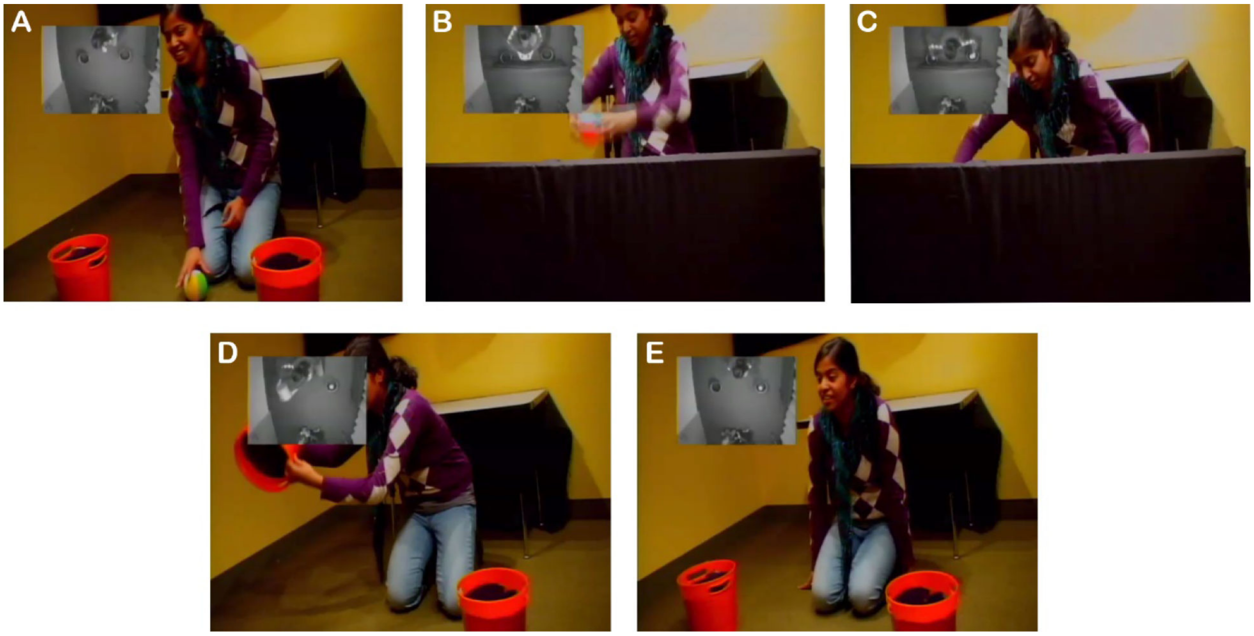

Figure 2.

Still photographs from a video of the experimental procedure used in test trials. (A) The experimenter shows the child that she is holding the ball, with the buckets fully visible. (B) The screen is placed in front of the buckets and the experimenter then lowers the ball at the midpoint between the two buckets, holding it with both hands while saying, “Look where it’s going!”. (C) The experimenter separates her hands behind the screen simultaneously while looking down at the floor, so the child cannot see which hand holds the ball. As she lowers the ball, she touches the bottoms of both buckets with her hands to mask any sound from the ball touching a bucket. (D) The screen is removed and the experimenter shows the child that one of the buckets is empty. (E) The experimenter asks, “Can you find the ball?”, inviting the child to search.

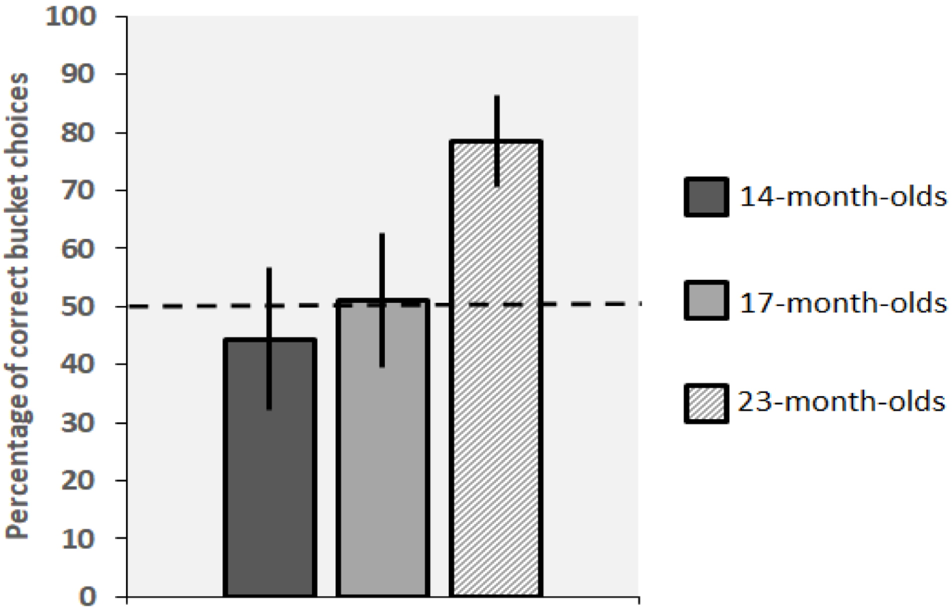

Figure 3.

Percentage of trials in which 14-, 17-, and 23-month-old infants approached the correct bucket in the two-cup exclusion task. Data from 14- and 17-month-old are from Experiment 1; data from 23-month-olds are from Mody & Carey (2016). Error bars represent 95% confidence intervals, and the dotted line indicates chance (50%).

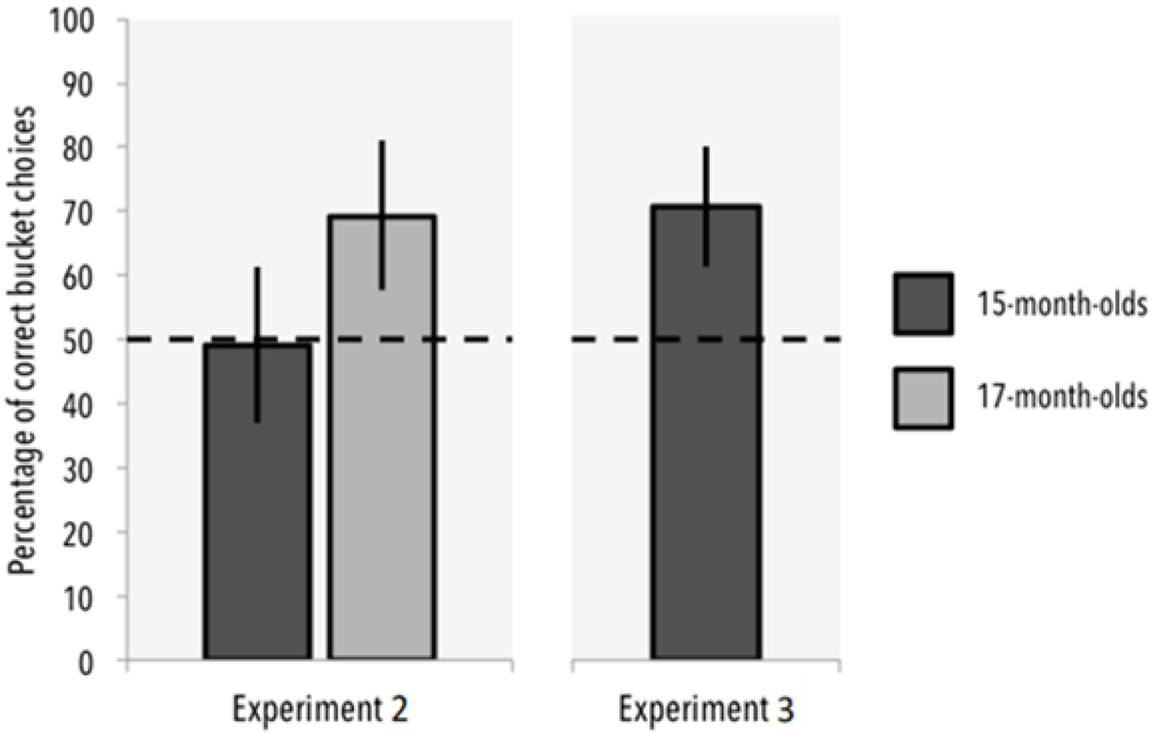

Figure 4.

Percentage of trials in which infants approached the correct bucket in Experiment 2 (two-bucket exclusion task, left) and Experiment 3 (two-bucket task with no inference, right). Error bars represent 95% confidence intervals, and the dotted line indicates chance (50%).

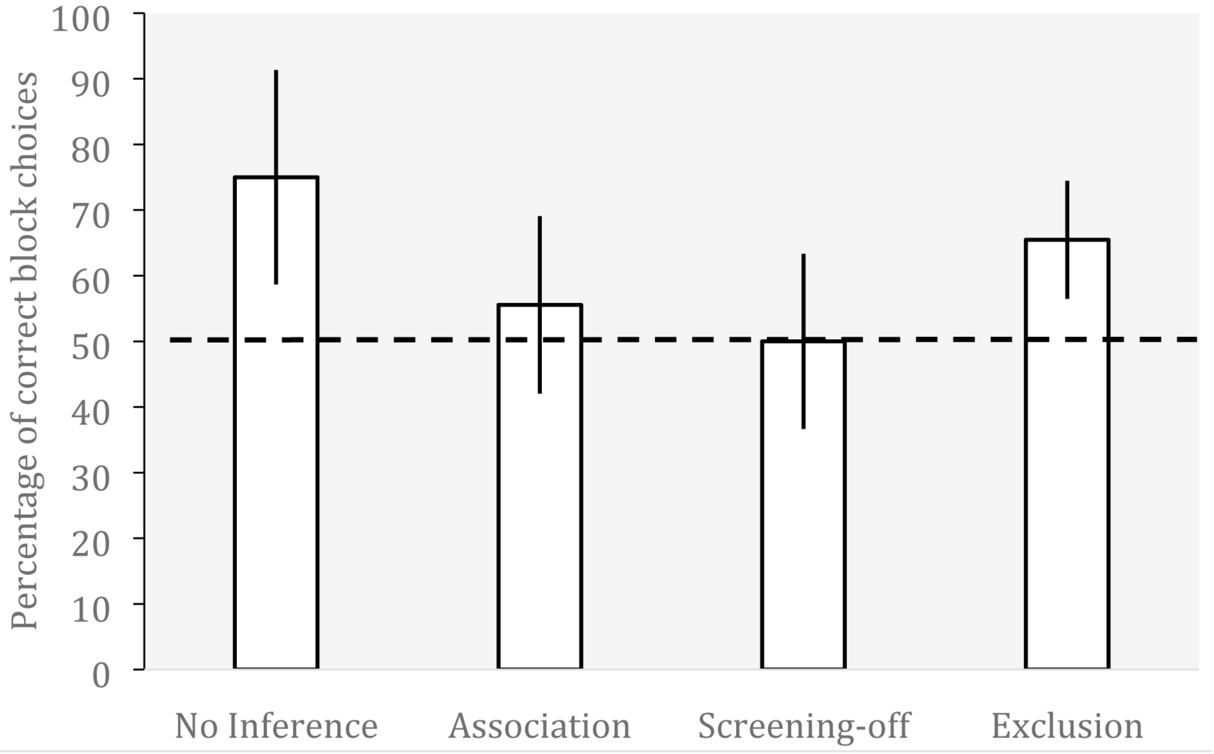

Figure 5.

Percentage of trials in which 19-month-old infants touched the active block to the blicket detector in each trial type of Experiment 3. Error bars represent 95% confidence intervals, and the dotted line indicates chance (50%).

Figure 6.

Schematic of the no-inference, exclusion, and association trials in Experiments 4 and 5. On no-inference trials, the experimenter activated the detector with one block, then failed to activate it with the other block. On exclusion trials, the experimenter activated the detector with both blocks twice, then failed to activate it with one of the blocks. On association trials, the experimenter activated the block with one block twice, and with the other block once. On all trials, the infant was then presented with both blocks and asked to “make it go.”

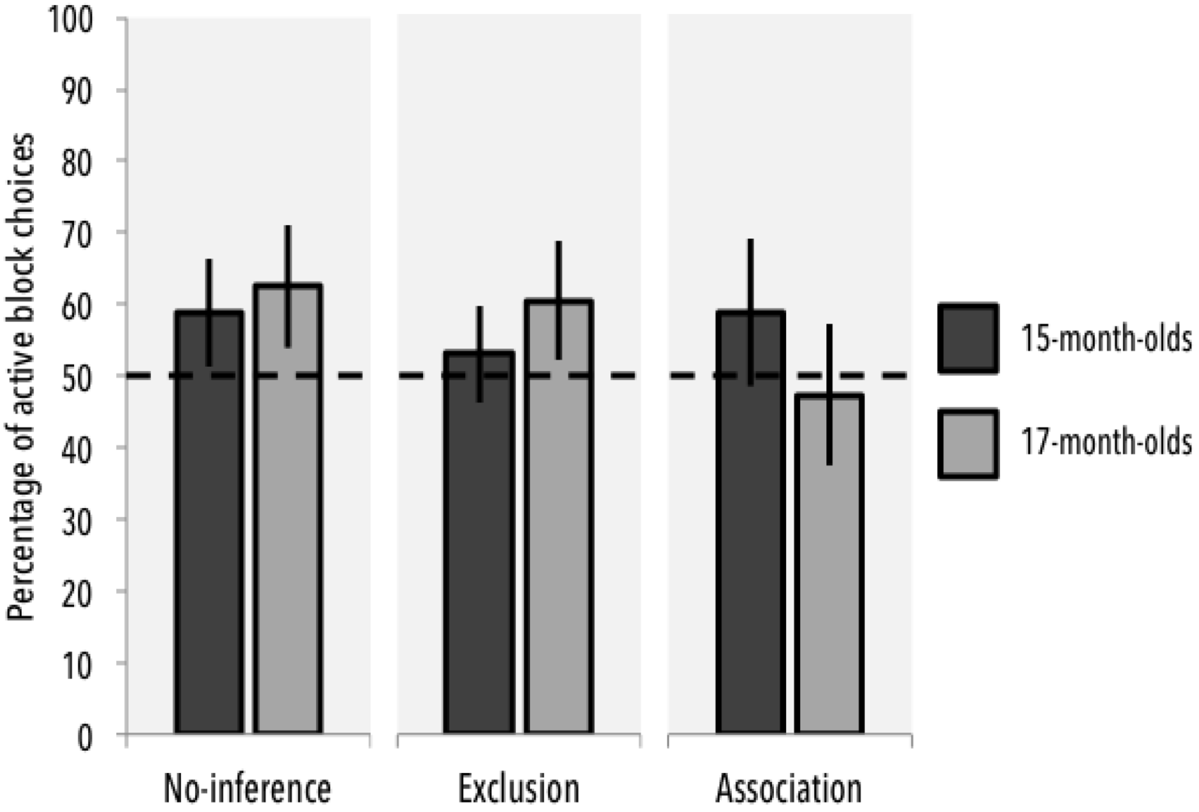

Figure 7.

Percentage of trials in which 15- and 17-month-old infants touched the active block to the blicket detector in each trial type of Experiment 5. For association trials, “active” refers to the block that had activated the detector twice rather than once. Error bars represent 95% confidence intervals, and the dotted line indicates chance (50%).