The Developmental Origins of Logical Inference:
Deduction and Domain-Generality

A dissertation presented
by
Shilpa Mody
to
The Department of Psychology

in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy
in the subject of
Psychology

Harvard University
Cambridge, Massachusetts

September 2016
The Developmental Origins of Logical Inference: Deduction and Domain-Generality

Abstract

Is there a fundamental divide between the types of thoughts that human adults can entertain and those available to infants and nonhuman animals? The research in this dissertation explores the developmental origins of abstract, combinatorial, propositional thought. As a case study, we examined infants’ and children’s ability to make a logical inference, the disjunctive syllogism: A or B, not A, therefore B.

In Paper 1, we asked when infants begin to recruit negation in reasoning. When shown that a toy was hidden in one of two buckets, and that one of those buckets was empty, 17-month-olds (but not 15-month-olds) used that negative information to preferentially approach the other bucket. When shown that two blocks placed on a toy together activated the toy, but that one of those blocks did not activate the toy by itself, 17-month-olds (but not 15-month-olds) used that negative information to preferentially attempt to activate the toy with the other block. This parallel onset, particularly when combined with a similar pattern of findings in the word learning literature, suggests that a common underlying factor led to the change in infants’ performance – likely, the ability to use negation flexibly in reasoning.

In Papers 2 and 3, we looked for evidence that children’s performance on these tasks actually reflected a deductive inference, rather than a simpler non-deductive strategy that did not require representing the disjunctive relation between the options.
particular, we looked for evidence that excluding one option led children to be deductively certain of the other option. In Paper 2, we used an adaption of the search task described above to assess whether seeing that one location was empty made children certain that the object was hidden in the other location. We found that 3-, 4-, and 5-year-olds showed evidence of making this inference, while 2.5-year-olds did not. Paper 3 used an analogous design in a version of the causal reasoning paradigm described above, and found that seeing that one block was inert led 3- and 3.5-year-olds to be certain that the other block could cause the toy to activate, while 2.5-year-olds showed no evidence of making this inference.

Together, these studies suggest that children begin to be able to recruit an abstract, combinatorial, domain-general representation of negation at around 1.5 years of age. However, it is not until 3 years of age that there is evidence that children actually make the disjunctive syllogism inference, recruiting representations of negation and disjunction to come to a deductively certain, logical conclusion.
# Table of Contents

**Introduction**
1. Animal & infant thought  
2. Why look at logic?  
3. Some definitions  
4. What we know about reasoning by exclusion  
5. Do these tasks reflect logical reasoning?

**Paper 1: Domain-General Negation in Infants' Reasoning**
1. Introduction  
2. Experiment 1 – Search Task  
3. Experiment 2 – Search Task with No Inference  
4. Experiment 3 – Causal Task  
5. General Discussion  
6. Supplemental Experiment S1  
7. Supplemental Experiment S2

**Paper 2: The Emergence of Reasoning by the Disjunctive Syllogism in Early Childhood**
1. Introduction  
2. Experiment 1  
3. Experiment 2  
4. General Discussion

**Paper 3: Young Children Make Logical Inferences in Causal Reasoning**
1. Introduction  
2. Methods  
3. Results  
4. Discussion

**Conclusion**
1. Summary of findings  
2. The disjunctive syllogism and Bayesian probabilistic inference  
3. More on negation  
4. More on disjunction

**References**
Acknowledgements

Working with Susan for the past six years has been a simply awesome experience. I get to have a weekly meeting with an incredibly smart, deep thinker who’s interested in all the same things as I am! Susan’s advising style strikes a perfect balance: making me feel empowered to become a my own independent mini-academic, while still feeling incredible well taken-care-of. If you know about attachment theory with infants, I’m basically talking about a case of really secure attachment. Thank you so very much.

To Jesse Snedeker and Liz Spelke – thank you tons and tons for many years of wonderful conversations and feedback – from helping me sort through ideas, to coming up with ideas for experiments, to just giving me external validation that my work is interesting (which was totally necessary at some stages). You’ve had a huge impact on both my work and my aspirations, and I really appreciate it.

To Steve Pinker and Josh Greene – we actually haven’t really talked about this work before. I’m pretty sure I’m supposed to be nervous about the scary closed-door question period after my defense… but I’m not, I’m just excited to have a chance to chat and hear your ideas. So, thanks in advance for that!

To Roman! What can I even say. We always joke about how we’re the same person, but the reality is even better than that: I think we complement each other really well. Spending all this time talking with you has made me a lot more thoughtful – about my work, my life, and of course my feelings. You’re the Gergely to my Csibra, and there is nobody I would rather be non-romantically spousal hired with.

To SHAME (my hilariously-named writing group) – this came as a surprise to me, but it turns out that I am groupy! Thanks for always being aggressive, whether it be celebrating victories or shaming failures, and for giving me 30 minutes a week to mostly just haggle about rules.

To Carey Lab and LDS – I have learned so much from you excellent people. From how to do good research, to how to ask good questions, to how to have a good time at conferences (or just hanging out in the lab office between babies), our community is the best.

To my awesome family and friends – my biggest hugs and warmest love. Everyone talks about how grad school is hard, and it was hard. But I have also been incredibly happy here, and that is due first and foremost to the amazing people I had the pleasure of surrounding myself with.

To Richard – I can’t imagine having a more supportive and loving partner than you. Thank you for following me to Cambridge, for being just as silly as I am (worried pug!), and most especially for always encouraging me to become a more awesome version of myself.
Introduction
1. Animal & infant thought

Human adults obviously have very different capabilities from those of infants or nonhuman animals. We are the only ones who organize global sporting events like the Olympics, engage in six years of research on toddlers’ logical reasoning abilities, or program apps to advise parents on the local average Tooth Fairy payout (adjusting for inflation)\(^1\). Our internal life also appears to be unique: we can ponder the possibility of life on a newly-discovered planet, the pros and cons of our presidential candidates, and what we plan to wear to our thesis defense. Is this difference only one of quantity, with our thoughts being incrementally more complex or variable than those of other creatures? Or do we think fundamentally different types of thoughts? In particular, do infants or animals have a Language of Thought – a combinatorial system for representing sententially-formatted, conceptual, propositional thoughts?

This question has been the target of debate among philosophers for centuries. Many thinkers, from Hume (1739/1978) to Davidson (1982) to Fodor (1975), have argued for continuity over ontogeny and phylogeny, while others, like Descartes (1637/1985) and Berwick and Chomsky (2016), have argued for a sharp divide, at least between animals and humans. These latter arguments have often hinged on the ability to use language, using animals’ lack of a human-like language as incontrovertible evidence that they cannot think as humans do; for example:

For it is quite remarkable that there are no men so dull-witted or stupid [...] that they are incapable of arranging various words together and forming an utterance from them in order to make their thoughts understood; whereas there is no other animal, however perfect and well-endowed it may be, that can do the like. [...] This shows not merely that the beasts have less reason than men, but that they have no reason at all: for [...] it would be incredible that a superior specimen of the monkey or parrot species should not be able to speak as well as the stupidest child [...] if their souls were not completely different in nature from ours (Descartes, 1637/1985, reproduced from Camp, 2009).

Yet this armchair assumption – that the only possible evidence for a propositional LoT is an external language – is untenable. Decades of research have shown that animals and infants have complex representational and computational abilities, ranging from the physical domain (e.g. Baillargeon, Spelke, & Wasserman, 1985; Herrmann, Call, Hernandez-Lloreda, Hare, & Tomasello, 2007; Wynn, 1992) to the social domain (e.g. Hare, Call, & Tomasello, 2001; 2006; Onishi & Baillargeon, 2005; Woodward, 1998). Yet if such complex thoughts can exist without language, can we reasonably argue that a propositional LoT necessarily requires an external language?

The possibility that animals and infants have a propositional LoT cannot be dismissed outright, or for that matter, assumed to be true. In addition to developing theoretical characterizations of the cognitive capabilities of animals and infants, this question must be informed by empirical evidence – do some representations of non-linguistic creatures have the properties of a propositional LoT?

2. **Why look at logic?**

The studies in this dissertation examine infants’ and young children’s ability to make a single, deductive, logical inference: the disjunctive syllogism – A or B, not A,
therefore B. We use this case study as a way to begin exploring the developmental origins of propositional thought, and the ability to use these thoughts for reasoning. The disjunctive syllogism is computed automatically by adults (Lea, 1995), and is one of the simplest and quickest inferences for adults to make (Braine, Reiser, & Rumain, 1984; Johnson-Laird, Byrne, & Schaeken, 1992; Rips, 1994), making it a plausible candidate for a logical inference that could be found early in development. But why look at logic?

2.1. Logical concepts are test cases for propositional thought

Logical operators, like “not” and “or”, are precisely the kinds of concepts that we would expect to find in a propositional LoT. They represent deeply abstract concepts; it is not at all clear how they could be represented iconically. Furthermore, they are deeply combinatorial: they can be combined into structured, rule-governed representations like “A or B” and “not A”, which can in turn be input into larger hierarchical structures or reasoning schemas. They are also semantically agnostic, and thus domain-general: a hallmark of logical inferences is that they are valid regardless of the specific content that they instantiate; they are based instead on the relationships between the constituents. This means that logical operators – unlike, for example, spatial or social relations – can be combined flexibly with truth-assessable content from any domain of knowledge.

One specific argument for taking logical concepts as markers of a propositional LoT comes from Tyler Burge (2010). He first differentiates propositional representations from perceptual ones (while noting that there are other forms of representation that do not fit into either of these categories, such as allocentric spatial maps, intermodal representations that guide action, or representations in core cognition). Perceptual representations always
refer to and are tied to specific things in the world that are being perceived, and attribute something to them – as in “that X thing”. Like perceptual attributives, conceptual attributives (the kind that can occur in propositional representations) are also able to make attributions to, and within the scope of, specific things in the world. So how can we tell them apart?

Burge argues that to pick out cases where a creature is thinking propositionally, we must look for instances of pure predication. Pure predication occurs when a (conceptual) attributive functions outside the scope of reference to a specific thing, and thus is not dependent on or bound to the identity of that specific thing – primarily, as in “that thing is X”. But again, how can we tell that a creature is thinking “that thing is X”, rather than “that X thing”? A useful feature of pure predication is that any conceptual attributive that is able to participate in such straightforward instances of pure predication can also occur in more complex thoughts, where – unlike a perceptual attributive – it might not actually make an attribution. He lays out three examples of such contexts:

1) When not making an attribution about any specific thing, such as in “every Y is X”
2) When not making an attribution itself, though it might be part of a larger attribution, such as in “that thing is not X” or “that thing is either X or Y”
3) When being part of something that makes no attribution at all, such as “it is not the case that that thing is X”

Finally, in addition to looking for instances of pure predication, Burge asserts (sadly, without explanation) that a second requirement for ascribing propositional thought to a creature is the ability to engage in propositional reasoning.
However, even apart from this specific condition, it is clear that many representations that would be used for propositional reasoning, particularly those involving negation and disjunction, would fall under the description of pure predication. For example, for an infant to represent a ball’s absence from one bucket (assuming she uses negation to do so), she could think “the ball is not in bucket A” or “it is not the case that the ball is in bucket A” – in neither case would “in bucket A” function to make an attribution – as in contexts 2 and 3 above – since the ball is truly not in bucket A. Thus, on Burge’s account, reasoning using disjunction and negation, such as by the disjunctive syllogism, is an ideal test case for whether infants and animals have access to a propositional LoT.

2.2. Logical concepts are test cases for other theories

Aside from a propositional LoT, numerous other hypotheses have been put forth to explain the difference between the cognitive abilities of animals (and sometimes infants) and human adults, several of which would also predict that the former creatures do not have access to logical concepts. For example, Bermudez (2003; 2006) proposes that animals cannot engage in logical inference – and in particular, the disjunctive syllogism – because representing and reasoning with logically structured thoughts requires that their truth-values and the relations between them be examined consciously and explicitly. He then goes on to argue that when we introspect, all of our second-order thoughts seem to be linguistically encoded, and thus language is the only plausible vehicle for logical inference.

Other theorists have suggested that animals and infants may be much more limited than adults in their ability to combine information flexibly; their thinking may be constrained to specific encapsulated domains of knowledge (Carruthers, 2002; Premack,
In other words, while animals and infants are capable of productive, abstract, and combinatorial thought within domains, they lack a domain-general language of thought. Since logical inferences and concepts are content-agnostic and domain-general, they should presumably not be available to infants and animals.

3. Some definitions

The general approach of the studies in this dissertation is to begin with tasks that appear, on their face, to recruit reasoning by the disjunctive syllogism: A or B, not A, therefore B. However, for many of these paradigms, there are numerous alternative explanations that might explain successful performance, without requiring domain-general logical concepts and deductive inference. But you have to start somewhere.

To describe successful performance on tasks that appear to – but might not – reflect inference by the disjunctive syllogism, I use the term reasoning by exclusion. This is meant to be agnostic as to the particular representations and inferential machinery that is being recruited. Successful reasoning by exclusion occurs when subjects are given two options and evidence that one of them is not correct, and they use this negative information to choose the other option.

In contrast, reasoning by the disjunctive syllogism or making the disjunctive syllogism inference requires that domain-general logical concepts of negation and disjunction are actually being used to think thoughts structured like “A or B” and “not A”, and that a deductive inference is made, resulting in a conclusion “therefore B”.
4. What we know about reasoning by exclusion

Research on children’s ability to reason using logical representations has focused largely on school-aged children, typically using tasks that rely heavily on language. For example, numerous studies looked at children’s ability to comprehend and make judgments about statements containing logical words like “and” and “or”, beginning around 7 years of age (e.g. Neimark & Slotnick, 1970; Paris, 1973). Others examined the inferences children make when presented with several of these statements together (Braine & Rumain, 1981; Osherson & Markman, 1975).

More recently, studies have focused on preschoolers’ comprehension of logical words, particularly “or”. These have revealed that children understand disjunction in a variety of linguistic contexts (e.g. Crain, Gualmini, & Meroni, 2000; Crain & Khlentzos, 2010; Chierchia et al., 2001; though see Singh, Wexler, Astle, Kamawar, & Fox, 2015 and Tieu et al., 2016) at least by the time they are 4 or 5 years old. And while studies of infants’ language production can be difficult to interpret\(^2\), it is worth noting that the word “no” is one of the earliest words that infants say, with parental reports indicating that about half of infants say “no” by 15 months of age (Dale & Fenson, 1996; Frank et al., in press). Whether

\(^2\) While “no” can clearly be used to express logical negation – as in, “there are no bears on Mars” – it can also be used for other functions: for example, to prohibit actions, reject offers, or express absence. Studies that categorize children’s production of negation have found that their early patterns of use are typically dominated by rejections, then followed by expressions of absence several months later; finally, after age 2, children begin to use negation to deny the truth of statements (Bloom, 1970; Pea, 1980, 1982). These earlier-emerging functions may or may not reflect a single, abstract concept of negation – at the very least, an infant who only uses “no” to reject foods she doesn’t like has not exhibited the flexibility in usage we might expect from logical negation.
this reflects access to an abstract concept of negation is an open question (see section 4.1 for evidence that children may not comprehend “no” until much later).

Apart from these investigations of their knowledge of logical words, little is known about whether very young children can represent and reason with logical representations. Yet while few studies have intentionally targeted logical representations like “or” and “not”, many use tasks that ask participants to reason by exclusion, and thus can be glossed as reflecting the disjunctive syllogism. These findings, in both young children and nonhuman animals, are summarized below.

4.1. Search

Call's (2004) cups task has been used extensively to test for reasoning by exclusion in the animal literature. In the cups task, an experimenter hides a reward in one of two cups. Subjects are then given evidence about the empty cup: they see or hear that it is empty. If they reason by exclusion, they should use the information about where the reward is not to exclude that location, and instead select the other cup.

Individuals of numerous animal species have been found to successfully reason by exclusion in this procedure, from great apes (Call, 2004; Call, 2006; Hill, Collier-Baker, & Suddendorf, 2011) to capuchin monkeys (Heimbauer, Antworth, & Owren, 2012; Paukner, Huntsberry, & Suomi, 2009; Sabbatini & Visalberghi, 2008) to African grey parrots (Mikolasch, Kotrschal, & Schloegl, 2011; Pepperberg, Koepke, Livingston, Girard, & Hartsfield, 2013; Schloegl, Schmidt, Boeckle, Weiß, & Kotrschal, 2012).

Yet the pattern of success is not uniform: among corvids, ravens, carrion crows, and Clark's nutcrackers succeed, yet jackdaws and Eurasian jays do not (Mikolasch, Kotrschal,
Dwarf goats succeed yet sheep do not (Nawroth, von Borrell, & Langbein, 2014). While most primates perform better in visual than auditory conditions, the only successful case among lemurs (a single individual!) was with auditory information (Maille & Roeder, 2012). Elephants succeed only when given olfactory information about which container is empty (Plotnik, Shaw, Brubaker, Tiller, & Clayton, 2014).

This cottage industry in testing various animals on the cups task has revealed not only differences across species, but important considerations of task design. Numerous studies have ruled out association- or learning-based accounts of the animals’ performance (e.g. Call, 2004; Hill et al., 2011; Pepperberg et al., 2013). Furthermore, numerous studies have attempted to rule out the possibility that animals are not actually representing the possible locations for the hidden food, but are simply acting based on aversion to the sight or sound of an empty cup (Call, 2006; Mikolasch et al., 2011; Pepperberg et al., 2013; Premack & Premack, 1994; see section 5.1.3 below).

In contrast to the animal work, only a few studies have examined children’s ability to reason by exclusion to find hidden objects. Three-, 4- and 5- year-old children readily solve the cups task, performing nearly at ceiling with both visual and auditory information (Hill, Collier-Baker, & Suddendorf, 2012). Furthermore, two recent studies used a variation on the cups task to ask whether young 2-year-old understand the words “no” and “not” in sentences that convey logical negation (Austin, Theakson, Lieven, & Tomasello, 2014; Feiman, Mody, Sanborn, & Carey, under review).

In the study by Austin and colleagues (2014), toddlers were shown that a block was hidden in either a toy house or a bucket, then given verbal information about where the
block was not (e.g. “Is it in the bucket?” “No” or “It’s not in the bucket”); with both sentence types, 24- and 28-month-olds searched in the correct container on about 80% of trials. In contrast, Austin and colleagues (2014) also tested a group of 21-month-olds, who chose between the two containers at chance. Using a similar paradigm in our lab, 20-month-olds actually performed below chance (Feiman et al., under review), choosing the bucket equally often when told “It’s not in the bucket” as “It’s in the bucket”. It is unclear whether these younger children’s performance was due to an inability to reason by exclusion, or a lack of understanding of the words “no” and “not”.

4.2. Causal reasoning

Many studies of causal reasoning in toddlers and preschoolers use an interactive “blicket detector” paradigm, where children are introduced to a toy that lights up when some, but not all, blocks are placed on it. They are then shown evidence about the causal status of individual blocks, and asked to make judgments about or choose among the blocks to intervene on the toy themselves. This paradigm has the benefit of allowing researchers to examine children’s causal reasoning about purely arbitrary causal connections, with the intervention measure also allowing a separation of children’s reasoning about causality from mere correlation or spatiotemporal association.

While no such studies have intentionally looked at children’s ability to reason by exclusion, several have used trials (I call these “exclusion” trials, though elsewhere they are termed “indirect screening-off” or “retrospective discounting”) that essentially ask children to do so (Sobel, Tenenbaum, & Gopnik, 2004; Sobel & Kirkham, 2006; Volter, Sentis, & Call, 2016). The earliest attested age of success on this task is 24 months: in this study, 19- and
24-month-olds saw that two blocks, A and B, activated the toy when placed on it simultaneously, and then that block A did not activate the machine by itself. When presented with both blocks and asked to activate the machine themselves, 24-month-olds preferentially put block B on the machine, while 19-month-olds chose between the two blocks at chance (Sobel & Kirkham, 2006). Additionally, success on this task is robustly found in 3-year-olds\(^3\) (Sobel, Tenenbaum, & Gopnik, 2004; Volter, Sentis, & Call, 2016).

The exclusion trial structure used in blicket detector studies actually originated in classical conditioning studies with animals, which were designed to investigate whether animals would exhibit “recovery from overshadowing”. Overshadowing refers to Pavlov’s (1927) classic finding that when a compound stimulus, say a simultaneously presented light and tone, is shown to predict an outcome, each individual stimulus acquires less associative strength than if it had been conditioned by itself. Some studies investigating the source of the overshadowing effect have followed overshadowing with an extinction phase, where one of the stimuli (e.g. the light) is presented without the corresponding effect – analogous to the negative information in reasoning by exclusion – and the animals’ response to the other stimulus (e.g. the tone) is measured. The results of this manipulation have been mixed, with some studies reporting that rats show decreased responding to the second stimulus (perhaps a kind of generalized extinction), and others reporting recovery from overshadowing – an increase in rats’ responding to the other stimulus, as if they

\(^3\) Volter, Sentis, and Call (2016) found that 2.5-year-olds’s performance on a similar task was indistinguishable from chance, though they suggest that this is likely due to their small sample size of only 15 children (I agree).
reasoned by exclusion (e.g. Matzel, Schachtman, & Miller, 1985; Holland, 1999; Burger, Hallemat, & Miller, 2000).

Excitingly, a single recent study extended the intervention-based bicket detector paradigm to great apes, again with mixed results. While apes performed at chance on typical exclusion trials (in two separate experiments), they succeeded on trials where they generated the evidence of the “premises” through their own actions, rather than by observation of an experimenter (Volter, Sentis, & Call, 2016). Further studies are needed to determine how robust this result is.

4.3. **Word learning & other cases of one-to-one mapping**

Reasoning by exclusion also appears to underlie the phenomenon known as “referent disambiguation” in word learning: when children hear a novel word in an ambiguous context, they resist mapping it onto objects for which they already know names, instead mapping it onto a novel object (e.g. Markman & Wachtel, 1988; Merriman & Bowman, 1989; Mervis & Bertrand, 1994; see Lewis & Frank, 2015, for a recent meta-analysis). There have been several proposals for the specific principle by which children reject the known object: for example, the mutual exclusivity constraint states that children assume that each object has only one name (Markman & Wachtel, 1988; Merriman & Bowman, 1989), while an account based on social pragmatics proposes that children assume that speakers will not be so uncooperative as to call an object by a name if it already has another one (Diesendruck & Markson, 2001). Regardless of the specific basis on which children reject the known object, most of these proposals have the same underlying structure: the known object is rejected, leading children to attend to the other
object instead – in other words, they involve reasoning by exclusion.

While many studies have focused on 2- and 3-year-olds, Halberda (2003) demonstrated that this ability is evident in even younger infants. In his study, 14- to -17-month-olds were shown two objects – one familiar and one novel, such as a ball and a phototube – on side-by-side computer screens. On trials in which they heard a known name ("Look at that ball! Ball!"), infants across the whole age range increased their looking to the named object, demonstrating that they knew its name. However, when they heard a novel name ("Look at that dax! Dax!"), only 17-month-olds increased their looking to the novel object, while younger infants did not. This success in toddlers beginning around 17 months of age has been replicated (Bion, Borovsky, & Fernald, 2012; Byers-Heinlein & Werker, 2009; White & Morgan, 2008) while a study with 16-month-olds produced mixed results (Markman, Wasow, & Hansen, 2003), suggesting that this ability comes online between 14 and 17 months of age.

Interestingly, two border collies also succeed at referent disambiguation (Kaminski, Call, & Fischer, 2004; Pilley & Reid, 2011). Both dogs had been taught hundreds of label-object associations, and apparently through this training learned that names and objects have a one-to-one mapping. When presented with several known-name objects and one novel object, and asked to fetch using a novel label, both dogs selected the novel object.

Several other tasks in the literature potentially reflect reasoning by exclusion in early childhood. Like the studies of word learning, these depend on children having an expectation of one-to-one mapping between two categories. For example, Moher, Feigenson, and Halberda (2010) trained 4- and 5-year-old children on mappings between faces and voices. When they heard a novel voice, they preferentially chose a novel face over
a face already associated with a voice. Using an analogous preferential looking paradigm, MacDonald and colleagues (2009) found that 2.5-year-olds increased looking to a novel animal upon hearing a novel animal vocalization.

5. Do these tasks reflect logical reasoning?

As stated above, the approach of the present work is to first find tasks that at least plausibly reflect use of the disjunctive syllogism. But what next? For each of the successes described above, there are other interpretations of children’s or animals’ performance that do not necessitate the full slate of logical resources that reasoning by the disjunctive syllogism demands. These vary from alternatives that do not use anything like negation, and thus cannot even really be described as reasoning by exclusion, to proposals that implement negation and a consideration of the relevant alternatives, but still fall short of deductive reasoning.

5.1. Alternatives that do not involve negation, and are not domain-general

5.1.1. Novel-novel matching

One alternative proposal for infants’ success at referent disambiguation is that children are positively drawn to apply novel names to novel (or name-unknown) objects, without specifically considering and rejecting known objects (Golinkoff, Mervis, & Hirsh-Pasek, 1994; Mervis & Bertrand, 1994). This principle does not require reasoning by exclusion, since children’s behavior is based on a positive inclination towards the novel object, rather than a negative one away from the known object. Analogous arguments could be made for the other cases that depend on one-to-one mapping – for example, children
might be positively inclined to map novel voices to novel faces.

However, there is evidence that 3- and 4-year-old children, as well as adults, do reason by exclusion in referent disambiguation paradigms: on trials where they already happen to be looking at the novel object when they hear a novel name, they show a characteristic pattern of first “double-checking” the known object as if to consider and reject it, before looking back towards the novel object (Halberda, 2006). Similar looking patterns may be found in 2.5-year-olds, across several one-to-one mapping tasks (Halberda, personal communication). While these findings are suggestive, it is not currently known whether they extend to infancy, or to nonhuman animals.

Crucially, an explanation based on novel-novel matching only applies to tasks that depend on one-to-one mapping – it cannot explain the successes in the search and causal reasoning tasks described above. If novel-novel matching underlies infants’ performance in referent disambiguation, there is no reason to expect commonalities in performance across it and the other tasks.

5.1.2. Contraries

Another possible account of success at any of the exclusion tasks described above is that infants or animals are using a representation that is constructed as a contrary to, rather than a negation of, a state of affairs. For example, infants might represent “the ball is absent from that bucket”, “that block is inert”, or “that thing is called ‘ball’”. On this proposal, infants do not need to be able to negate “that bucket contains the ball”, they only need to know that it is contrary to “the ball is absent from that bucket” – that is, that the two states cannot coexist (Bermudez, 2006).
Importantly, contrary descriptions are not constructed using a general-purpose formal rule that employs a common negation operator, but are based on specific semantic knowledge about which pairs of states cannot coexist (e.g. absence vs. presence, being causally inert vs. being causally active, having the name ‘dax’ vs. having the name ‘ball’). This means there is nothing in common between the descriptions “the ball is absent from that bucket”, “that block is inert”, and “that thing is only called ‘ball’”. Thus, the use of contrary descriptions would not predict any commonalities in infants’ or animals’ performance across contexts.

5.1.3. Avoidance

Numerous studies of the cups task in nonhuman animals have forwarded an alternative proposal for the animals’ success: perhaps they are simply avoiding the empty cup, and end up searching in the correct cup merely as a default, since it is the only other thing to do (e.g. Paukner, Huntsberry, & Suomi, 2009; Mikolasch et al., 2012). On this interpretation, animals are maintaining no representation of the hidden food in memory, and thus are not considering the two cups as alternative possibilities in any meaningful way. Since they do not represent the hidden food, they cannot be thinking “the food is not in that cup”. They simply find the sight of an empty cup, or the shaking of a cup accompanied by silence, aversive.

To test for this possibility, several studies have used a design in which two different foods are hidden in the two containers, in view of the subject. The experimenter then removes one of the foods, allowing the animal to see which food was removed but not which container it was removed from, and discards or eats it. The animal has not
experienced the aversive sight or sound of an empty cup, and must instead determine where to search based on a representation in memory of which food was hidden in which cup. This paradigm has revealed substantial individual differences in ability, but at least some chimpanzees, parrots, and 4-year-old children reliably succeed, indicating that they made their choice based on representations of the foods’ locations (Call, 2006; Mikolasch et al., 2011; Pepperberg et al., 2013; Premack & Premack, 1994).

This design has not been used with infants, nor have any studies investigated whether infants are simply avoiding an inert block in the blicket detector paradigm. Thus we do not currently know whether their success on these tasks depends on a representation of the possible locations of the hidden object or the possible causes of the blicket detector’s activation, or if it is simply based on avoidance. However, an avoidance explanation is less tenable for referent disambiguation – why would infants find the known object aversive? Furthermore, as in the explanation based on contraries, there is no principled commonality between finding inert blocks aversive and finding empty buckets aversive – since there is no common representation or operation involved, if infants reasoned by exclusion by avoidance, there would be no reason to expect commonalities in their performance across these tasks.

5.2. Are infants using negation to reason by exclusion? (Paper 1)

For any given success at reasoning by exclusion in infancy, there are open alternative explanations that do not require negation. Infants could succeed at referent disambiguation by novel-novel mapping, and the causal and search tasks by avoidance of the incorrect alternative. They could succeed at all three by acting on a representation that
is contrary to, rather than a negation of, a positive representation of a state of affairs.

In contrast, the proposal that infants really are deploying a general-purpose negation operation on these tasks, and therefore reasoning by exclusion, makes a unique prediction: there might be commonalities in infants’ performance across the tasks. That is, despite differences in background knowledge required, how performance is measured, and executive function demands, infants may begin to succeed at these tasks around the same time – when they are able to deploy negation to reason by exclusion.

In Paper 1, we explore 15- and 17-month-old infants’ access to a flexible, domain-general concept of negation by examining their ability to reason by exclusion across different contexts. In Experiments 1 and 2, a toy is hidden in one of two buckets, then one bucket is revealed to be empty, and we ask children to search for the toy. In Experiment 3, we show infants that two blocks placed on a toy together activate it, then that one of those blocks does not activate the toy by itself, and ask infants to activate the toy. We look for evidence of parallel emergence in the ability to reason by exclusion for searching, causal reasoning, and – based on evidence from the literature – referent disambiguation.

Of course, even if performance on these tasks reflects negation, infants could be limited from showing their true competence by other factors; to the extent that these other factors vary across the tasks, infants’ reasoning abilities might be uncovered differentially. For example, infants might lack the specific knowledge that names and kinds are mapped one-to-one, leading to a delay on referent disambiguation despite an ability to reason by exclusion. Yet while the interpretation of differences across tasks would be ambiguous, a convergence across tasks is most parsimoniously explained by a common factor driving that convergence – likely, infants’ ability to flexibly deploy negation in reasoning.
5.3. **An alternative that does not involve disjunction, and is not deductive**

5.3.1. *Independent possibilities*

Even if we are convinced that infants use negation to reason by exclusion, this does not get us all the way to the full disjunctive syllogism. One alternative explanation, which we put forth in Papers 2 and 3, is that infants and animals could successfully negate one possibility without changing their assessment of the other possibility, because they have failed to represent the disjunctive “or” relation between them.

For example, in the cups task, children may represent the possible locations of the toy as bucket A and bucket B. However, they could do so without making any commitment to those possibilities being exhaustive, or even to them being related. Rather than representing that the toy is in either A or B, they could represent that the toy might be in A, and, independently, that it might be in B. Upon seeing that A is empty, they could negate that possibility, yet fail to update their assessment of B – they would then search in B because the toy *might* be there. This explanation does not require disjunction, since the possibilities are represented independently. It also does not result in deductive certainty in the conclusion – in fact, it does not result in any increase in certainty at all.

While this alternative has not been explicitly stated in the literature, three studies have looked for the evidence that would counter it: if reasoning “not A” results in increased certainty in B, A and B must have been represented relationally. These studies have found evidence that reasoning by exclusion results in increased certainty in search contexts in 4- to 6-year-old children – but, crucially, not in 2.5-year-old children, and not in either dogs or chimpanzees (Call & Carpenter, 2001; Watson et al., 2001). The only study of causal
reasoning found evidence for increased certainty in 3.5-year-olds, the only group tested (Beckers, Vandorpe, Debeys, & de Houwer, 2009)

For example, Call & Carpenter (2001) tested chimpanzees and 2.5-year-old children on a task where a reward was hidden inside one of three opaque tubes. Before subjects selected a tube, they could look inside the tubes to see if the reward was hidden there. On trials where subjects happened to look inside the two empty tubes first, they could then use the disjunctive syllogism to deduce that the third tube must contain the reward before looking in it. However, the chimpanzees chose the third tube without looking in it on only 14% of such trials, while the 2.5-year-olds did so on 5% of such trials. This low rate of selection is consistent with the proposal that seeing that A and B were empty did not increase children’s or chimpanzees’ certainty that the reward was in C. Instead, they acted as if they represented only the initial premise “maybe C”, checking to see if it contained the reward before selecting it.

5.4. Are children using disjunction to make a deductive inference? (Papers 2 and 3)

An explanation of reasoning by exclusion based on independently-represented possibilities proposes that children might know what the relevant possibilities are (A and B), and correctly exclude one based on negative evidence about a constituent (not A), yet their actions might still be based on a representation about which they are not deductively certain. In fact, they might act based on an assessment of B that has not changed at all – the initial premise, “maybe B”.

In Papers 2 and 3, we look for evidence that children’s assessment of B increases after seeing “not A”. In Paper 2, we adapt the cups task to ask if seeing that one cup is
empty leads 2.5- to 5-year-old children to the certain, deductive conclusion that the cup
paired with it contains a reward. In Paper 3, we ask the same question of 2.5-, 3- and 3.5-
year-old children in a causal reasoning paradigm. This evidence can rule out the possibility
that children have represented the possibilities independently; instead, such evidence
would indicate that children represent the disjunctive relation between possibilities, and
combine this logically structured representation with negative information to yield a novel,
deductively certain conclusion. If children inferentially update their assessment of B based
on information about A, it would demonstrate both that they are using disjunction, and that
they are engaging in deductive inference.

Furthermore, if children begin to show signs of recruiting disjunction in reasoning
at the same time across several different tasks, we can rule out still more possible
alternative explanations. In particular, this would make it unlikely that children are
succeeding based on domain- or task-specific “copies” of the disjunctive syllogism – for
example, a version of the disjunctive syllogism that only specifies how to search or forage,
or one that only implements a word learning constraint like mutual exclusivity. While the
possibility of multiple domain-specific copies would not be ruled out definitively, a more
parsimonious explanation would be that the ability to recruit disjunction for reasoning, or
to combine disjunctively-structured information with negation in deductive inference,
emerges at this time, driving the onset of success across multiple tasks.
Paper 1: Domain-General Negation in Infants’ Reasoning

Shilpa Mody, Roman Feiman, & Susan Carey
1. Introduction

Negation is one of the most basic concepts in thought, but is also among the most abstract and expressively powerful. It gets this expressive power from being highly flexible: negation can be combined with any propositional thought, regardless of its contents, to flip its truth-value. Here we ask whether infants show signs of having access to such a flexible, domain-general concept of negation, by examining their ability to reason by exclusion in several different contexts. Given two options, and information that one of them is incorrect, do they use that negative information to choose the other option?

Beginning around 17 months, infants appear to reason by exclusion in the context of word learning (Bion, Borovsky, & Fernald, 2012; Byers-Heinlein & Werker, 2009; Halberda, 2003; White & Morgan, 2008). For example, in one “referent disambiguation” study (Halberda, 2003), 14- to -17-month-olds saw two objects – one novel, and one whose name the child knew – displayed on side-by-side computer screens. On trials where they heard a known name (“Look at that ball! Ball!”), all infants increased their looking to the named object. However, when they heard a novel name (“Look at that dax! Dax!”), only 17-month-olds increased their looking to the novel object, while younger infants did not. Most analyses of this behavior propose that children are attempting to disambiguate the meaning of the new word, and do so by excluding the known object as a potential referent (Diesendruck & Markson, 2001; Markman & Wachtel, 1988; Merriman & Bowman, 1989),

---

4 This study reported that 18-month-olds performed at chance on referent disambiguation trials, but a footnote reveals that they did increase looking to the novel object above baseline. Due to a baseline preference to look at the known object, this increase did not result in >50% looking to the novel object.
leading children to attend to the other object instead. In other words, they involve reasoning by exclusion.

However, referent disambiguation might not reflect an ability to deploy negation in thought. First, infants might be positively drawn to apply novel names to novel objects, without considering and rejecting known objects - they might not excluding at all (Golinkoff, Mervis, & Hirsh-Pasek, 1994; Mervis & Bertrand, 1994). Second, even if referent disambiguation uses exclusion, it might reflect a domain-specific word-learning process, rather than a flexible, domain-general concept of negation. To determine whether infants can deploy negation in reasoning more widely, we need converging evidence from tasks that cannot be solved by matching novelty to novelty, and that are unrelated to word-learning. Accordingly, in this paper, we ask when infants’ ability to reason by exclusion extends to two very different situations: searching for a hidden object, and inferring the cause of an event.

Reasoning by exclusion in a search context has been widely studied in nonhuman animals using the cups task (Call, 2004). Chimpanzees, gorillas, and bonobos saw an experimenter hide a piece of food in one of two covered cups. On critical trials, they were shown the inside of the empty cup, or it was shaken and made no sound, and the apes could then select one of the cups. If they reasoned by exclusion, they should use the information about where the reward was not to exclude that location, and instead select the other cup. The apes performed nearly perfectly given visual information, and were also above chance given auditory information. The list of species in which at least some individuals succeed now stretches from capuchin monkeys (Sabatini & Visalberghi, 2008) to African grey parrots (Pepperberg et al., 2013; Schloegl et al., 2012) to elephants (Plotnik, Shaw,
Brubaker, Tiller, & Clayton, 2014), among many others. Looking at children, 3-, 4-, and 5-year-olds robustly succeed in both auditory and visual conditions (Hill, Collier-Baker, & Suddendorf, 2012), and 2-year-olds succeed when they are verbally told which container does not hold the ball (Austin, Theakson, Lieven, & Tomasello, 2014; Feiman, Mody, Sanborn, & Carey, under review). We recently conducted two studies with even younger children: 23-month-olds and 20-month-olds consistently approached the correct location upon being shown that the other bucket was empty (Feiman et al., under review; Mody & Carey, 2016).

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, are shown evidence about the causal status of individual blocks, then are asked to make judgments about or choose among the blocks. While no studies have focused on children’s use of negative information, several have used trials (dubbed “indirect screening-off”) that ask children to reason by exclusion (Sobel, Tenenbaum, & Gopnik, 2004; Sobel & Kirkham, 2006). 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 reasoned by exclusion, rejecting the inert block and attempting to activate the toy using the other block. In an initial study we found that 19-month-olds also succeed (see Supplemental Experiment S2).

The present experiments explore when the ability to reason by exclusion emerges in development. In Experiments 1 and 2, we assess infants’ ability to use negative information
about an empty location to direct their search to another location. In Experiment 3, we examine whether they can make the same inference regarding causality. We ask whether infants’ ability to reason by exclusion in these two tasks converges with the literature on referent disambiguation in word learning, showing a parallel shift between 14 and 17 months. Such a finding would suggest that the three tasks draw on a common, domain-general ability to reason using negative information that emerges at this age.

2. **Experiment 1 – Search Task**

2.1. **Methods**

2.1.1. **Participants**

We tested two groups of infants: 15-month-olds ($N = 24$, $M_{age} = 15.02$ months, range = 14.07-16.01, 12 boys) and 17-month-olds ($N = 24$, $M_{age} = 17.62$ months, range = 17.00-18.48, 15 boys). We selected this sample size, which is in the typical range used in our lab (e.g. Feiman et al., under review; Mody & Carey, 2016), before testing began. We found a large effect in 23-month-olds using a very similar procedure (Cohen’s $d = 1.30$; see Supplemental Experiment S1), suggesting that this sample size was sufficient; 90% power for detecting an effect of this size in a one-sample t-test required only 7 participants.

Participants were recruited by phone and email and were tested at the Laboratory for Developmental Studies at Harvard University. Children were given a small gift and parents were compensated $5.00 for travel expenses. 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).

2.1.2. Materials & procedure

The stimuli consisted of four pairs of cloth-lined buckets and a large black screen. Each trial used two identical buckets; the color of the buckets varied across trials to reduce perseveration. We asked caregivers to pick which of three small toys (ball, rubber duck, stuffed dog) their child would be most interested in finding.

Infants were held on their caregiver’s lap, who sat on the floor 6’ away from the experimenter. Caregivers were asked to close their eyes while the toy was being hidden and the empty bucket was revealed. Each child participated in two warm-up trials and eight test trials. Some children were 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.

**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 toy above the bucket, called for the child’s attention, and then lowered the toy into the bucket with both hands in full view. She then asked the child to find the toy. On the second warm-up trial, the experimenter placed the screen in front of the bucket, lowered the toy into it, then removed the screen and asked the child to find the toy. 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 warm up trial.

**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
toy 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 toy, 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 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.

We coded children’s responses 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 toy until they approached a bucket or approximately 10 seconds elapsed. If they did not find the toy themself, the experimenter showed them where it was; regardless of their actions, at the end of the trial children were given 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.

2.2. Results

We terminated the study after four or six test trials for five 17-month-olds and four 15-month-olds. 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).

The results of Experiments 1 and 2 are summarized in Figure 1. We found that 17-month-olds approached the correct bucket on 69.3% of trials, which was greater than chance, $t(23) = 4.569, p < .001$, Cohen’s $d = .93$. Analyzing the 19 infants who completed at least 7/8 trials, there was no evidence of a learning effect between the first four test trials (65.8% correct) and the last four test trials (71.5% correct), $t(18) = 0.79, p = .440, d = .18$. Furthermore, 17-month-olds were marginally successful on the first trial, sign test, $17/24$

![Figure 1](image)

**Figure 1**

Percentage of trials in which 15- and 17-month-old infants approached the correct bucket in Experiment 1 (search task, left), and percentage of in which 15-month-old infants approached the correct bucket in Experiment 2 (search task with no inference, right). Error bars represent 95% confidence intervals, and the dotted line indicates chance (50%).
correct choices, \( p = .064 \), suggesting that they reason by exclusion spontaneously. In contrast, 15-month-olds approached the correct bucket on only 49.2\% of trials, which was indistinguishable from chance, \( t(23) = 0.145, p = .886, d = .04 \). 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 (49.6\% correct) and the last four test trials (56.3\% correct), \( t(19) = 1.05, p = .307, d = .04 \). Comparing the two age groups, 17-month-olds approached the correct bucket significantly more often than 15-month-olds, \( t(46) = 3.2, p = .002, d = .93 \). Strikingly, this pattern of results mirrors the emergence of referent disambiguation in word learning.

The pattern of results was very similar when looking at only the first four test trials, to facilitate comparison across experiments. Seventeen-month-olds chose the bucket that contained the toy on 70.8\% of trials, which was better than chance, \( t(23) = 4.995, p < .001, d = 1.02 \). However, 15-month-olds chose the correct bucket on only 46.5\% of trials, which was at chance, \( t(23) = 0.615, p = .545, d = .13 \). Seventeen-month-olds’ performance was again significantly better than 15-month-olds’, \( t(46) = 3.47, p = .001, d = 1.00 \).

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 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 49.0\% of trials, while the remaining 17-month-olds were at 81.0\%.

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, this 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. At neither age did infants demonstrate a tendency to approach whichever bucket was manipulated last.

While there was no effect of the order that the buckets were manipulated in, it was important to call children’s attention to both buckets rather than only the empty one. We ran an earlier study in which the experimenter showed children that one bucket was empty, but left the other bucket untouched. When the pragmatic cues highlighting both choices were not equated, both 15- and 17-month-old infants chose at chance levels (see Supplemental Experiment S1).

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-attended bucket, leaving open the possibility that it was due to a real inability to reason by exclusion. However, other 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; these possibilities are addressed in Experiment 2.
3. **Experiment 2 – Search Task with No Inference**

3.1. **Methods**

3.1.1. **Participants**

The participants were 24 15-month-olds ($M_{age} = 15.00$ months, range = 14.10-15.98, 9 boys). We kept the same sample size as in Experiment 1, which was selected before testing began. Recruitment and compensation were identical to Experiment 1. An additional 9 toddlers participated, but were excluded from the final sample for failure to search on warm-up trials (8) or parental interference (1).

3.1.2. **Procedure**

We asked whether 15-month-olds were both able and motivated to find a hidden toy under conditions that were very similar to Experiment 1, but did not require reasoning by exclusion. We repeated the procedure with two modifications. First, the hiding procedure allowed children to see where the toy was being hidden. 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 into the two buckets; infants could see which bucket the toy was heading for. She then removed the screen, manipulated both buckets as in Experiment 1, and asked the child to find the toy. Second, since almost 20% of children in Experiment 1 had become fussy over the course of the eight test trials, each child participated in only four test trials.

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.

3.2. Results

For six 15-month-olds, 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.

When they knew where the toy was hidden, 15-month-olds approached the correct bucket on 70.5% of trials, which was greater than chance, \( t(23) = 4.280, p < .001, \) Cohen’s \( d = .87 \). 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 (70.8%) and the last two test trials (72.9%), \( t(23) = 0.33, p = .744, d = .07 \). Despite the overall high rate of success, we found evidence of perseveration: seven 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: \( p = .048 \)).

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 2 performed better than those in Experiment 1, \( t(46) = 3.24, p = .002, d = .94 \), indicating that 15-month-olds’ failure to reason by exclusion was not due to being distracted by the bucket manipulations or their limited memory or motivation. Instead, it is likely that their failure in Experiment 1 was due to an inability to use the information that one bucket was empty to direct their search away from that bucket. While 15-month-olds can use positive information about where a toy is to direct their searching, they do not appear able to use negative information about where it is not.
The results of Experiments 1 and 2 show that 17-month-olds, but not 15-month-olds, can reason by exclusion when searching for a hidden toy; this timeline mirrors previous results from referent disambiguation. In Experiment 3, we asked whether the shift in performance between 15 and 17 months would be mirrored in a third context: reasoning by exclusion to determine the cause of an event.

4. Experiment 3 – Causal Task

4.1. Methods

4.1.1. Participants

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). Since the blicket detector method is not commonly used with infants this young, we decided to increase our standard sample size by 50% prior to beginning testing. We found a medium-sized effect in 19-month-olds using a very similar procedure (Cohen’s d = 0.60, see Supplemental Experiment S2), suggesting that this sample size was sufficient; 90% power for detecting an effect of this size in a one-sample t-test required 30 participants. Recruitment and compensation were identical to the previous experiments. An additional five 15-month-olds and four 17-month-olds were tested but excluded from the final sample for completing fewer than seven out of ten trials with usable data (8) or parental interference (1). A further seven 15-month-olds were excluded for making no response on the first three trials, which led us to abort testing.
4.1.2. Materials & 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 blocks that were identical in shape but differed in color. 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 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 used.

Each child participated in 10 trials: one warm-up trial, three no-inference trials, four
exclusion trials, and two association trials; the trials were always presented in this order. The color and side of the active blocks varied across trials and were counterbalanced across children. The structure of each trial type is summarized in Figure 2.

**Warm-up trial.** Children were introduced to the blicket detector with four blocks in front of it. The experimenter demonstrated each block in succession: the first and fourth blocks activated the toy, while the second and third did not. She 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.

**No-inference trials.** No-inference trials gave infants direct information that one block was active and another was inert. These trials were 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 asked the child to make it go.

**Exclusion trials.** These trials required that infants reason by exclusion to infer which block to select. 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. If infants reasoned by exclusion, they should use the information that this block is inert to exclude it, and try the other block instead.

**Association trials.** These trials tested whether children were choosing blocks based on a simple associative strategy. On the first association trial, the experimenter activated
Figure 2
Schematic of the no-inference, exclusion, and association trials in Experiment 3. 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”.

38
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.

4.2. 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 22 trials across 11 15-month-olds and two 17-month-olds due to the child touching both blocks to the blicket detector simultaneously; while this response is rational – it maximizes the chance of activating the detector – it is also difficult to interpret. The results of Experiment 3 are summarized in Figure 3.

On no-inference trials, we found that 17-month-olds tried to activate the blicket detector using the active block on 62.5% of trials, which was greater than chance, $t(35) = 2.87, p = .007$, Cohen’s $d = .48$. Similarly, 15-month-olds chose the active block on 58.8% of trials, which was greater than chance, $t(35) = 2.26, p = .030$, $d = .38$, and not significantly different from the 17-month-olds, $t(70) = 0.64$, $p = .524$, $d = .15$. Infants at both ages were motivated to activate the blicket detector, made choices based on the causal information they were shown, and could switch their attention away from the (inert) block that was manipulated last.

However, only 17-month-olds succeeded at exclusion trials. They tried to activate the blicket detector using the active block on 60.2% of trials, which was greater than
chance, $t(35) = 2.40, p = .022, d = .40$. There was no evidence of a learning effect between the first two exclusion trials (58.3%) and the last two exclusion trials (62.9%), $t(34) = 0.68, p = .501, d = .12$ (one infant generated no usable data on the last two exclusion trials and was dropped from this comparison).

In contrast, 15-month-olds chose the active block on only 52.8% of trials, which was indistinguishable from chance, $t(35) = 0.80, p = .424, d = .40$. There was no evidence of a learning effect between the first two exclusion trials (48.6%) and the last two exclusion trials (56.9%), $t(35) = 1.43, p = .162, d = .24$. However, the difference in accuracy between

Figure 3

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 3. 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%).
the two groups was not significant, $t(70) = 1.36, p = .178, d = .32$. While these findings are not quite as robust, the overall pattern of results converges with the emergence of both referent disambiguation and, as explored in Experiments 1 and 2, reasoning by exclusion in a search context.

Finally, we found no evidence of any consistent preference on association trials: 17-month-olds chose the block that activated the blicket detector twice on 47.1% of trials, $t(34) = 0.57, p = .571, d = .10$, and 15-month-olds chose it on 58.6% of trials, $t(34) = 1.64, p = .110, d = .28$ (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.

We then considered several alternative explanations for the 15-month-olds’ chance performance on exclusion trials. Based on their success on no-inference trials, their exclusion performance could not 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 a difficulty with attention-switching could explain 15-month-olds’ chance performance on exclusion trials.

A final possibility is that the younger infants perseverated in their responses over the course of the study. 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 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 17-month-olds chose the active block on 62.5% of no-inference trials, \(t(31) = 2.55, p = .016, d = .45\), and 61.5% of exclusion trials, \(t(31) = 2.42, p = .022, d = .43\). The remaining 15-month-olds chose the active block on 62.1% of no-inference trials, \(t(21) = 2.26, p = .035, d = .48\), but only 54.5% of exclusion trials, \(t(21) = 0.81, p = .426, d = .18\).

5. General Discussion

Experiments 1 and 2 investigated infants’ ability to reason by exclusion when searching: they saw a toy being hidden in one of two buckets, then that one bucket was empty; we asked whether they would use this negative information and search in the other bucket. We found that 17-month-olds approached the correct bucket, while 15-month-olds searched at chance, and across the two experiments we ruled out numerous performance demands as potential reasons for their failure. Experiment 3 asked whether infants could reason the same way about causality. We showed infants that two blocks together activated a toy, but that one of them alone did not, and asked whether they would attempt to activate the toy using the other block. Again, we found that 17-month-olds succeeded and that 15-month-olds were at chance, and we ruled out several performance demands that could have limited younger infants’ behavior. This suggests that the ability to reason by exclusion to find a hidden object or to infer the cause of an event emerges only around 1½ years old.

What might account for younger infants’ failure? Studies of the A-not-B error have shown that infants may fail to correctly search for an object due to memory interference across trials or a weak representation of the object (Munakata, 1998; Smith & Thelen, 2003). These studies were conducted with much younger infants, but it is possible that
similar failures of memory or executive function led to the 15-month-olds’ at-chance behavior in our more complicated search paradigm. The blicket detector task could have been difficult for younger infants due to the ambiguous information conveyed by showing two blocks together activating the toy. However, neither of these explanations account for a striking aspect of the data: the transitions from 15 to 17 months are the same across three contexts: search, causality, and word-learning. This parallel suggests that a common underlying change might be responsible for the shift in infants’ performance.

Critically, the three paradigms are all very different. They depend on different background knowledge and use different measures. Referent disambiguation could plausibly be based on a positive inclination to map novel words to novel objects, rather than reasoning by exclusion, but no similar alternative applies to the other tasks. The paradigms also make different demands on executive function. The search and blicket detector tasks require holding newly presented negative information in working memory, while referent disambiguation does not. All three tasks require switching attention endogenously away from the negative object or location, but a general failure of attention-switching cannot explain the 15-month-olds’ behavior: in Experiment 3, they successfully switched away from the inert block on no-inference trials, yet still failed on inference trials. Furthermore, on half the trials in Experiment 1, we actively drew infants’ attention away from the empty bucket after revealing its contents, and 15-month-olds still failed.

Despite the differences between the tasks, infants begin to succeed on them at the same age. Of course, this could be a coincidence, with improvements in task-specific knowledge or executive function happening to occur around the same time, allowing infants to reveal latent reasoning abilities. Or the parallel could be principled – 15-month-
olds’ may not be able to use a domain-general concept of negation to reason by exclusion. This could be due to the processing demands inherent in this type of reasoning – for example, reasoning on the basis of a thought that is itself composed of two meaningful pieces (negation and the thought being negated). Alternatively, younger infants may struggle with negation itself – they may not be able to deploy it flexibly in thought, or even have access to the concept at all. In either case, future studies examining individual differences in, or priming across, these tasks could shed light on whether the relationship between them is real or coincidental, and whether it is independent of differences in children’s executive function.

When 17-month-olds successfully reason by exclusion, what might their concept of negation be like? One possibility is that they deploy a symbol for negation, in the sense of an adult-like, truth-functional operator that can combine freely with other information (e.g. to form thoughts like “the ball is not in that bucket”, “that block does not make it go”, or “‘dax’ does not refer to the ball”). However, a more limited negation procedure could also do the job: upon seeing that one bucket is empty or that one block is inert, or determining that the ball is not the referent of “dax”, the child might eliminate that possibility from consideration. An option-elimination procedure could be domain-general, supporting reasoning by exclusion in many contexts, and clearly implements one function of conceptual negation. However, such a procedure could not be used to think thoughts where negation is embedded within other functions (e.g. “A or not B”, “if not A then B”, conditioning a response on “not A”, etc.). One hint that infants may not have ready access to symbolic negation is that they have surprising difficulty learning a word for it: despite hearing “no” very frequently, and even saying it themselves well before 17 months of age
(Dale & Fenson, 1996), children do not appear to understand the truth-functional meaning of “no” and “not” until after their second birthday (Austin et al., 2014; Feiman et al., under review).

The present studies showed that infants begin to reason by exclusion to find a hidden object and to infer the cause of an event between 15 and 17 months, the same age that they begin to succeed at referent disambiguation. If this is not simply a coincidence, the parallel emergence of these abilities likely hinges on what they have in common: reasoning on the basis of negative information. These findings suggest that a domain-general concept of negation is not easily deployed in infants’ reasoning until 1½ years of age, and leave open the possibility that it is not part of their conceptual repertoire at all.

6. Supplemental Experiment S1

6.1. Methods

6.1.1. Participants

We tested 17-month-olds ($N = 24$, $M_{age} = 17.70$ months, range = 17.06-18.51, 11 boys) and 14-month-olds ($N = 24$, $M_{age} = 14.41$ months, range = 13.55-15.35, 13 boys). An additional 11 14-month-olds and 13 17-month-olds were excluded for not searching on warm-up trials (17), parental interference (3), completing fewer than three usable test trials (4), or experimenter error (2). The analyses also include data from a previously tested group of 23-month-olds (see Mody & Carey, 2016).

6.1.2. Procedure.

The procedure was nearly identical to Experiment 1 (for details see Mody & Carey,
The main difference was that the experimenter manipulated only the empty bucket: she lifted it and showed children that it was empty, but did not draw attention to the other bucket.

6.2. Results

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). Results are summarized in Figure 4.

**Figure 4**

Percentage of trials in which 15-, 17-, and 23-month-old infants approached the correct bucket in Experiment S1. Error bars represent 95% confidence intervals, and the dotted line indicates chance (50%).
Seventeen-month-olds approached the correct bucket on 51.0% of trials, which did not differ from chance, $t(23) = 0.176, p = .862$, Cohen's $d = .04$. Similarly, 14-month-olds searched in the correct bucket on 44.4% of trials, which did not differ from chance, $t(23) = 0.891, p = .382, d = .18$. We found no evidence that this 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: $p = 0.705$; 17-month-olds: $p = 1.000$). In contrast, the 23-month-olds approached the correct bucket on 78.5% of trials, which was greater than chance, $t(23) = 6.362, p < .001, d = 1.30$.

Combining all the data, a one-way ANOVA found a main effect of age, $F(2, 69) = 10.412, p < .001, \eta^2 = .23$. Post-hoc Bonferroni tests revealed that 23-month-olds performed better than 17-month-olds ($p = .003, d = 1.07$) and 14-month-olds ($p < .001, d = 1.28$), who did not differ from each other ($p = 1.000, d = .22$).

### 6.2.1. Comparison to Experiment 1

The 17-month-olds who participated in Experiment S1 performed significantly worse than those in Experiment 1, $t(46) = 2.73, p = .009, d = .79$. In contrast, there was no difference between the 14- and 15-month-olds across these experiments, $t(46) = 0.25, p = .804, d = .07$. The main change to the procedure was that the empty bucket was always manipulated just before children were asked to search, and it was the only bucket the experimenter attended to. Since we found that infants had no predisposition to approach the last-manipulated bucket in Experiment 1, the most likely explanation for 17-month-olds' failure in Experiment S1 is that they were biased to approach the only bucket that the experimenter highlighted with pragmatic cues.
7. Supplemental Experiment S2

7.1. Methods

7.1.1. Participants

The participants were 19-month-old infants ($N = 28, M_{age} = 19.79$ months, range = 18.71-20.58, 14 boys). We intended to test 24 infants, but accidentally continued testing. An additional 11 infants were excluded for completing fewer than seven out of ten trials with usable data (9), crying (1), or failing to act on the first three trials (2), which led us to abort testing.

7.1.2. Procedure

Each child participated in one warm-up trial, one no-inference trial, two association trials, two screening-off trials (not used in Experiment 3), and four exclusion trials. The procedure for each trial was identical to Experiment 3, except for changes detailed below.

**Warm-up trial.** Between the demonstration of the second and third blocks, the experimenter demonstrated the first and second blocks on the toy together twice, activating it.

**No-inference trial.** The experimenter repeated the demonstration sequence, thereby demonstrating each block twice.

**Screening-off trials.** 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.
7.2. 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). Results are summarized in Figure 5.

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 S2. Error bars represent 95% confidence intervals, and the dotted line indicates chance (50%).

On the no-inference trial, 21 out of 28 infants tried to activate the detector using the active block, which was greater than chance, sign test, $p = .013$. Yet surprisingly, on screening-off trials – which also 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, $t(27) = 0.00$, $p = 1.000$, $d = 0$.

On exclusion trials, infants chose the active block on 65.1% of trials, which was greater than chance, $t(27) = 3.23$, $p = .013$, $d = .60$. Furthermore, 21 out of 28 infants selected the active block on their first exclusion trial, sign test, $p = .013$.

Finally, on association trials, infants chose the block that had activated the detector twice on 55.8% of trials, which was no different than chance, $t(26) = 1.142$, $p = .264$, $d = .23$; there was no indication that infants were using a simple associative strategy to make their selection.

Unlike the results reported by Sobel & 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. However, 19-month-olds’ surprising failure at screening-off trials led us to simplify the procedure when we tested younger infants in Experiment 3: we dropped the screening-off trials, and included more no-inference trials.
Paper 2: The Emergence of Reasoning by the Disjunctive Syllogism in Early Childhood

Shilpa Mody & Susan Carey
1. Introduction

Philosophers and cognitive scientists have long debated whether nonhuman animals and prelinguistic infants have a language of thought that is qualitatively different from our own. One difference that has been proposed is that animals and infants may be much more limited in their ability to combine information flexibly or to think abstract, combinatorial thoughts (e.g. Carruthers, 2002; Penn, Holyoke, & Povinelli, 2008; Premack, 2007; Spelke, 2002). On this hypothesis, animals and infants may lack the ability to represent logical concepts like the or and not of classical logic, think logically structured thoughts like “A or B”, and make deductive inferences like “A or B, not A, therefore B”. Logical concepts are deeply combinatorial – they represent nothing but the relationship between other constituents of thought. They are also deeply abstract – the hallmark of logical inferences is that they are valid regardless of the specific content that they instantiate. Logical inferences and the representations that make them up are therefore strong candidates for being represented in an abstract, combinatorial language of thought that animals and infants may not possess.5

5 The question of whether infants and animals have the capacity for logical inference is independent of the theoretically important questions of how these inferences are represented and computed. See, for instance, the debates between mental model theorists (e.g. Johnson-Laird, 2010) and natural deduction system theorists (e.g. Braine, 1978; O’Brien, Braine, & Yang, 1994; Rips, 1994). Here we are concerned with the orthogonal question of when logical capacities emerge in ontogeny. For simplicity, we primarily use the language of natural deduction in this paper, but our proposal applies equally to a mental model conception. On this latter story, the difficulty that children may face in reasoning by the disjunctive syllogism is not in manipulating or combining propositional thoughts, but in implementing those thoughts into mental models that are properly structured and
In this paper, we focus on one simple logical inference, the disjunctive syllogism: A or B, not A, therefore B. The disjunctive syllogism requires representing a disjunctive or between two possible states of affairs: either one or the other is true. When one possibility is ruled out with not, this information can be combined with the disjunction to generate novel information: the other possibility must be true. To make this inference, it is necessary to represent – or at least implement – the concepts or and not as defined in classical, propositional logic. While adults can clearly make inferences that are beyond the scope of classical logic – for example, those involving quantifiers, modal operators, graded probabilities, or degrees of belief – concepts with the meanings of classical logic’s or and not are ubiquitous in adult human’s thought. The disjunctive syllogism is computed automatically by adults (Lea, 1995), and is one of the simplest and quickest inferences for adults to make (Braine, Reiser, & Rumain, 1984; Johnson-Laird, Byrne, & Schaeken, 1992; Rips, 1994). These considerations make it a good candidate for a case study of abstract, combinatorial thought in non-linguistic animals and prelinguistic infants.

One additional reason for choosing this case study is that many studies in the literature on animal cognition have already begun to address it. One task that potentially reflects reasoning by the disjunctive syllogism is Call’s (2004) cups task, which has been used to test for what is called “reasoning by exclusion” in the animal literature. In the cups task, an experimenter hides a reward in one of two cups. On critical trials, subjects are then given evidence about the empty cup: they see or hear that it is empty. If they reason by exclusion, they should use the information about where the reward is not to exclude that evaluated – this is sketched out in further footnotes.
location, and instead select the other cup. Individuals of numerous animal species have been found to successfully reason by exclusion in this procedure, including great apes (Call, 2004; Hill, Collier-Baker, & Suddendorf, 2011), siamangs (Hill et al., 2011), olive baboons (Schmitt & Fischer, 2009), capuchin monkeys (Heimbauer, Antworth, & Owren 2012; Paukner, Huntsberry, & Suomi, 2009; Sabbatini & Visalberghi, 2008), lemurs (Maille & Roeder, 2012), dogs (Erdohegyi, Topal, Viranyi, & Miklosi, 2007), ravens (Schloegl et al., 2009), carrion crows (Mikolasch, Kotrschal, & Schloegl, 2012), and African grey parrots (Pepperberg et al., 2013; Schloegl et al., 2012). Three-, 4- and 5-year-old children also readily solve the cups task (Hill, Collier-Baker, & Suddendorf, 2012), and 24- and 27-month-olds succeed on a version in which the information is conveyed verbally (Austin, Theakson, Lieven, & Tomasello, 2014).

But is these animals’ and children’s success on the cups task evidence for an ability to use the disjunctive syllogism? While the term “reasoning by exclusion” does not make any commitments as to the particular reasoning mechanism being used, many have suggested that success on the cups task reflects logical inference-making, referring to it as “causal-logical inference” (Call, 2004) or “inferential reasoning by exclusion” (Hill et al., 2011). In particular, success is often ascribed to working though a disjunctive syllogism: when the subject sees or hears that one cup is empty, they infer that the reward must be in the other cup (e.g. Schmitt & Fischer, 2009). However, success on the cups task is open to at least three interpretations that differ substantially in their required logical and representational properties.

On the richest interpretation, subjects are truly implementing a disjunctive syllogism: A or B, not A, therefore B. This requires a representation of the dependent
relationship between A or B that is embodied in or: one of A or B must be true, so information about A affects the subject’s appraisal of B. When they see that A is empty, it leads them to update their representation of B, and they conclude that B necessarily contains the reward. This interpretation requires representing the concepts or and not. It also requires substantial combinatorial ability: composing two logically structured thoughts “A or B” and “not A” and then combining those to generate the new information “B”6.

However, another interpretation of success on the cups task is that subjects do not set up the initial premise “A or B”, but instead consider the two possible hiding locations independently – we call this the “maybe A, maybe B” interpretation. They represent A and B as the two possible locations of the reward, but do not represent the dependent relationship between them. When they see that A is empty, they remove it as a possible location and avoid searching in it, but since the information about A and B was represented independently, this does not lead them to update their appraisal of B. Rather than concluding that B necessarily contains the reward, they search in B based on their initial premise that it might contain the reward. This “maybe A, maybe B” interpretation requires some way of implementing the thought “not A” – we speculate about how this might be

---

6 The disjunctive syllogism interpretation could be articulated in mental models by the animal initially representing the location of the reward with two mental models: one model where the reward is in A, and another where the reward is in B. When A is eliminated, only the second model remains, so it is evaluated as true. The two models are connected by or, since as a set of alternatives they exhaust the space of possibilities under consideration, and they are mutually exclusive; this disjunction is not explicitly symbolized, but is implicit in how the models are established and evaluated.
achieved in the General Discussion. But importantly, the “maybe A, maybe B” interpretation certainly does not require representing the logical concept OR. It also has fewer combinatorial demands than a full logical inference, since the thought “A OR B” is never composed, and there is no new conclusion generated.

A third interpretation is that subjects do not have an ongoing representation of the alternatives A and B as potential locations of the hidden reward. When they see that A is empty, they avoid searching in it, and instead approach B merely because it is the other salient hiding location available to them. In this “avoid empty” interpretation, subjects have no particular beliefs about whether B contains the reward; they are merely not searching in A. Again, while this interpretation does depend on some way of implementing “NOT A”, it does not require representing the logical concept OR, and has few combinatorial demands.

The cups task is ambiguous in regards to these three alternatives; using any

7 The “maybe A, maybe B” interpretation could be articulated in mental models by the animal initially constructing two separate models: one where the reward is in A, and another where the reward is in B. However, these models are not considered together as a set of alternatives; instead, they are generated by separate premises (“maybe A” and “maybe B”). Each model is marked as uncertain by pairing it with an implicit “…” model such as those used to describe conditionals (e.g. Johnson-Laird et al., 1992). When A is eliminated, the remaining set of models constitutes an uncertain representation of the reward in B (“maybe B”), upon which the animal bases its search. Another possible articulation is that the animal initially has a single model that ambiguously represents the reward as possibly in A and possibly in B. When A is eliminated, that part of the model is updated, but the representation of the possibility that the food is in B is unaffected. Thus, the animal looks in B because the food might be in B (“maybe B”).

8 The “avoid empty” interpretation could be articulated in mental models by the animal initially having no model of the reward’s location at all. When shown that A is empty, the animal avoids A, and merely searches somewhere else.
approach would lead to success. Furthermore, all three of these interpretations involve reasoning: they require some representation of the environment, and result in generating new information. They could all lead to rational and efficient behavior – namely, effective reasoning by exclusion. However, only the disjunctive syllogism involves representing the logical concept OR and combining logically structured thoughts, and is thus clearly an example of abstract, combinatorial thought.

Previous attempts to pinpoint the mechanism behind the cups task have focused on ruling out the leanest “avoid empty” interpretation, establishing that at least some animals do have working memory representations of the food that has been hidden. For example, in one design, subjects are shown two different foods being hidden in the two cups in full view. Next, the containers are hidden, and the experimenter removes one of the foods – the subject cannot see which container it is being removed from – and shows it to the subject. The animals are then allowed to approach the containers. Some individual apes and parrots reliably succeed at this task by selecting the cup with the remaining food (Call, 2006; Mikolasch, Kotrschal, & Schloegl, 2011; Pepperberg et al., 2013; Premack & Premack, 1994). To do so, they had to have inferred which cup was empty, demonstrating that they had some representation of the hidden foods, rather than having simply learned a behavioral rule “avoid cups that are empty”. These studies demonstrate that at least some animals do have working memory representations of the location of hidden foods in this choice context, and use this information to pick a location that might contain a reward. However, they do not address whether the animals actually use the logical concept OR to represent the dependent relationship between those possible locations, and in doing so infer where the hidden food must necessarily be.
If subjects are using the concepts OR and NOT to make the logical inference – and not representing the locations independently or simply avoiding the empty location – they should update their belief about one possibility when finding out that the other possibility is not true. In particular, they should conclude that the reward is necessarily in the other location; this notion of necessity is central to deductive inference. This means that to adjudicate between the logical account on the one hand, and the two non-deductive accounts on the other hand, we can look for behavioral evidence that subjects are engaging in inferential updating, inferring that the reward is certain to be in one location upon seeing that it is not in the other. Only two previous studies have directly tested for this inferential updating signature of deductive inference in nonhuman animals and young children. Both studies failed to find compelling evidence for its use in animals, while the results for children were varied: 4- to 6-year-olds showed evidence of inferential updating, while 2.5-year-olds did not.

Call & Carpenter (2001) tested chimpanzees and 2.5-year-old children on a task where a reward was hidden inside one of three opaque tubes. Before subjects selected a tube, they could look inside the tubes to see if the reward was hidden there. On trials where subjects happened to look inside the two empty tubes first, they could then use the disjunctive syllogism to inferentially update their appraisal of the third tube, concluding that it must contain the reward before looking in it. However, the chimpanzees chose the third tube without looking in it on only 14% of such trials, while the 2.5-year-olds did so on 5% of such trials. In another experiment with only two tubes, the rate at which apes stopped looking and made a choice was roughly the same whether the first tube was empty or blocked (such that it provided no information about whether it contained the reward).
The apes’ equivalent behavior in the two conditions suggests that seeing the empty tube did not lead to any change in their expectations of the remaining tube. In these experiments, neither chimpanzees nor toddlers appeared to spontaneously deploy the disjunctive syllogism; their behavior was consistent with the non-deductive strategies.

The second study showed domestic dogs and 4- to 6-year-old children an experimenter walking behind three screens, depositing a reward behind one of them (Watson et al., 2001). They were then asked to search for the reward. After searching behind the first two screens and finding them empty, they could use the disjunctive syllogism to infer that the reward was necessarily behind the third one. Plausibly, this increase in certainty would correspond to an increase in subjects’ speed as they moved from the second to the third screen, compared to when they had moved from the first to the second screen. This was observed for the children, suggesting that they inferred that the third screen must hide the reward. In contrast, the dogs’ speed decreased as they moved to the third screen, an effect that is consistent with extinction. There was no evidence that they inferentially updated their belief about the third location based on reasoning through a disjunctive syllogism; again, the dogs’ behavior was consistent with the non-deductive strategies.

In these studies, all the subjects successfully reasoned by exclusion: when they encountered an empty hiding location, they tended to move their search to another location. But while 4- to 6-year-olds demonstrated evidence of inferential updating, neither chimpanzees, dogs, nor 2.5-year-olds showed evidence of having made the disjunctive syllogism inference. Rather, the extant evidence is equally compatible with these populations using one of the non-deductive strategies.
In the current study, we further examine the emergence of the representation of the logical concepts OR and NOT in children, in the context of reasoning by the disjunctive syllogism. In Experiment 1, we establish whether 23-month-old toddlers are able to spontaneously reason by exclusion in Call’s cups task. Since some animal species require extensive training through numerous trials to pass the cups task (e.g. Maille & Roeder, 2012), and in many non-human species only a subset of individuals succeed, it is plausible that even reasoning by exclusion, regardless of the mechanism that underlies it, might not be in the repertoire of very young children. Having established that children under 3 years old are able to reason by exclusion, in Experiment 2, we use a novel variant of the cups task to disambiguate which mechanism 2.5- to 5-year-olds use to do so: do they simply avoid the empty cup as they continue searching (“avoid empty”) or pick any cup that might contain a reward (“maybe A, maybe B”), or does seeing that one location is empty lead them to inferentially update their belief about the other location?

2. **Experiment 1**

Three-, 4-, and 5-year-old children succeed at the cups task (Hill et al., 2012), while 2-year-olds succeed at a variant in which the information about which location is empty is conveyed verbally (Austin et al., 2014). Experiment 1 asks whether 23-month-olds also spontaneously reason by exclusion.

2.1. **Participants**

The participants were 24 23-month-old toddlers (mean age = 23.6 months, range = 23.0-24.0, 13 boys). This sample size was chosen before testing began. Participants were
recruited by phone and email and were tested at the Laboratory for Developmental Studies at Harvard University. Children were given a small gift and parents were compensated $5.00 for travel expenses. One additional toddler participated but was excluded from the final sample due to failure to search for the ball on warm-up trials.

2.2. Methods

The stimuli consisted of four pairs of cloth-lined buckets, a ball, and a large black screen. Each trial used two same-colored buckets, and the color of the buckets varied across trials to reduce perseveration.

Toddlers were held on their caregiver’s lap, who sat on the floor approximately 6’ away from the experimenter. Caregivers were asked to close their eyes while the ball was being hidden and the empty bucket was revealed, but could watch while their children searched in the buckets. Each child participated in two warm-up trials and four test trials.

2.2.1. Warm-up trials

Each session started with two warm-up trials using only one bucket, designed to familiarize the child to the task and apparatus. The first warm-up trial did not use the screen. The experimenter lowered the ball into the bucket with both hands in full view, then immediately asked the child to find it. On the second warm-up trial, she placed the screen in front of the bucket, lowered the ball into the bucket with both hands, removed the screen, and then asked the child to find the ball.

If children failed to search for the ball on the first warm-up trial, they proceeded to the second warm-up trial. If children failed to search on the second warm-up trial, they
were given a third identical warm-up trial. To proceed to the test trials, children had to search in the bucket on at least one warm up trial (one participant failed to do so).

2.2.2. **Test trials**

On each of the four test trials, the experimenter placed two identical buckets in front of herself, each equidistant from midline and 38” from each other. She then covered the buckets with the screen and held the ball above the center of the screen. The screen fully covered the bucket from the child and caregiver’s view, but still allowed them to see the experimenter’s upper body and face. She caught the child’s attention and lowered the ball with both hands. When her hands were behind the screen, she separated them and lowered each hand into a bucket, so that the child could not see where she was hiding the ball. She removed her hands and showed the child that they were empty. After removing the screen, the experimenter demonstrated that one of the buckets was empty by turning it upside down, shaking it, showing the child the inside of the bucket, and then placing it back in its original position. She then asked the child to find the ball, keeping her eyes on the child. The caregiver released the child, who was free to approach one of the buckets.

If the child did not approach one of the buckets within 5 seconds, the experimenter encouraged them to search for the ball until they approached one of the buckets or approximately 10 seconds elapsed. If they did not approach a bucket within 10 seconds, or approached the incorrect bucket, the experimenter showed them where the ball was. Two orders for the location of the ball were constructed – (left, right, right, left) and (right, left, left, right) – and each order was used for half the children.
2.3. Results

For each of seven children, one test trial was excluded due to the caregiver releasing them before the empty bucket had been revealed (6) or the child’s failure to approach one of the buckets (1). Using the remaining test trials, a score was computed for each child.

The toddlers approached the correct bucket on 78.5% of trials, which is significantly greater than chance ($t(23) = 6.362, p < .001$). There was no evidence of a learning effect between the first two test trials (77.1% correct) and the last two test trials (79.2% correct) ($t(23) = 0.238, p = .812$). Furthermore, the toddlers were marginally successful on their first trial (sign test, 17/24 correct choices, $p = .064$), suggesting that they reason by exclusion spontaneously and without training.

These results establish that 23-month-olds, like 3- to 5 year-olds, can reason by exclusion in the cups task – when they see that one of two locations is empty, they direct their searching to the other location. However, as discussed above, success on this task can be explained by three different reasoning processes: the disjunctive syllogism, and two non-deductive strategies that do not require any representation of the concept OR and do not result in inferential updating. Experiment 2 distinguishes between the disjunctive syllogism and the two non-deductive accounts.

3. Experiment 2

In this task, children competed with a second experimenter to find stickers that were hidden inside four cups. Two stickers were hidden inside the four cups, one sticker in each pair of cups (Figure 6). One cup was then revealed to be empty. If children were reasoning using the disjunctive syllogism, they could combine this information (NOT A) with their
representation of where the sticker was hidden (A or B) to conclude that the cup paired with the empty cup necessarily contained a sticker (THEREFORE B), while the location of the other sticker was unsure. In this case, we would expect children to preferentially choose the cup paired with the empty cup (hereafter, the certain choice is called the “target cup”).

Figure 6
Structure of training trials (left) and test trials (right) in Experiment 2. Symbols above the cups indicate the information that is available to the participant: the cup with the cross is empty, the cup with the checkmark is certain to contain a sticker, and the cups with question marks may or may not contain stickers.
However, if they were using the “maybe A, maybe B” strategy, learning that one cup was empty (NOT A) would eliminate that cup as a potential location for a sticker, but would not lead to updating information about the target cup (MAYBE B); all three remaining cups would be equally good candidates for containing a sticker. In this case, we could expect children to choose the target cup at an equal rate as the other two cups.

Finally, if they were using the avoid-empty strategy, learning that one cup was empty (NOT A) would lead them to avoid that cup, but they would have no other representation about the possible locations of the stickers. Thus, we would again expect children to choose the target cup at an equal rate as the other two cups, since they are all equally salient possibilities.

3.1. Participants

The participants were 96 children: 24 2.5-year-olds (mean age = 2.8 years, range = 2.5-3.0, 12 boys), 24 3-year-olds (mean age = 3.5, range = 3.0-3.8, 12 boys), 24 4-year-olds (mean age = 4.5, range = 4.0-5.0, 12 boys), and 24 5-year-olds (mean age = 5.6, range = 5.1-6.2, 15 boys). This sample size for each age group was chosen before testing began. Most participants were recruited and tested at the Boston Children’s Museum; some were recruited by phone and tested in the lab. Children tested in the lab were given a small gift and their parents compensated $5.00 for travel expenses. In addition, two 2.5-year-olds and two 3-year-olds were excluded from the final sample due to failure to complete the study (3) or interference by the caregiver (1).
3.2. Methods

Four paper cups were used as hiding locations; they were covered with different colored paper to increase their distinctiveness. A small white screen was used to conceal the cups during hiding. A variety of small stickers were used as rewards; two identical stickers were used in each trial.

Experimenter 1 sat across a table from the child and Experimenter 2. The cups were arranged in front of Experimenter 1 with two cups to the left and two cups to the right. The distance between each cup in a pair was approximately 4", and the distance between the pairs was approximately 12".

At the start of the session, Experimenter 1 explained that the child and Experimenter 2 would take turns to pick cups and win the stickers inside them. Importantly, she also explained that each trial would end after the first sticker was found, to motivate children to pick correctly on their first choice. Each child participated in a training phase (three training trials intermixed with two demonstration trials), followed by a test phase (four test trials intermixed with one filler trial).

On all trials, if the child chose a cup that contained a sticker, they were given the sticker, and the trial ended. If they chose a cup that did not contain a sticker, Experimenter 1 showed them that the cup was empty and asked Experimenter 2 to choose a cup; Experimenter 2 always chose the target cup, which ended the trial. In addition, if children did not immediately respond on any trial, they were encouraged to make a choice until they responded or it was clear that they were unwilling to respond.
### 3.2.1. Training phase

The three training trials (Figure 6) involved only three cups: one pair of cups and one single cup (the target cup). The cup that was not used varied across the trials. On each trial, Experimenter 1 first covered the left set of cups with the screen, held a sticker above the center of the screen in both hands, and then lowered the sticker behind the screen. If there was one cup behind the screen, she dropped the sticker into the cup with both hands. If there were two cups behind the screen, she put each hand into a cup while dropping the sticker, so the child could not distinguish which cup it was put in. The screen was then moved to the right set of cups and the hiding procedure was repeated. Finally, the screen was removed, and the child was asked to choose a cup. Thus, on training trials children were faced with a choice between three cups: one that certainly contained a sticker and two that may or may not have contained a sticker. Training trials, like test trials, required children to compare the sure cup to the two uncertain cups, but did not involve reasoning by exclusion.

Two demonstration trials were included during the training phase: the order of trials in this phase was demonstration, training, demonstration, training, training. The sticker hiding events in demonstration trials were identical to those in training trials. However, after the stickers were hidden, Experimenter 2 was asked to choose a cup instead of the child, and she always chose the target cup, ending the trial. While choosing, she explained her reasoning, e.g. “On this side, the sticker could be in the red cup or in the green cup, I’m not sure. But on this side, the sticker must be in the blue cup, so I’m going to choose the blue cup”. This was done to cue children to pick the cup that necessarily contained a sticker instead of guessing, since this was required for a correct response on
both training and test trials.

The location of the stickers was counterbalanced across trials; the same order of trials was used for each child.

3.2.2. Test phase

The four test trials (Figure 6) were identical to training trials, with the following exceptions. First, all four cups were used. Second, after the stickers were hidden in the cups, Experimenter 2 was asked to choose a cup; Experimenter 2 always chose an empty cup. After this was shown to be empty, the child was asked to choose a cup. The child was thus presented with the same choice as in training trials: two cups that might or might not contain a sticker, and one cup that necessarily contained a sticker. Unlike training trials, however, they had to arrive at this certainty by reasoning through the disjunctive syllogism.

Across the test trials, which of each pair of cups contained stickers was counterbalanced across children, as was the cup that Experimenter 2 selected. For each child, each cup contained a sticker on two of the test trials, and each cup was selected once by Experimenter 2.

One filler trial was included in the test phase; the order of trials in this phase was test, test, filler, test, test. After the stickers were hidden in the cups, the child was asked to choose a cup first. Since they had no information about which cups were empty, they could only respond by guessing.
3.3. **Results**

For seven of the 96 children, one test trial was excluded due to the child’s unwillingness to respond. For two children, one training trial was excluded for the same reason. Using the remaining trials, the percentages of target cup choices were computed for each child (Figure 7).

3.3.1. **Training trials**

We established chance at 33%, as there were three cups to choose among. Since there were four age groups being compared to chance, we used the Bonferroni correction for multiple comparisons, leading to an adjusted alpha of .0125. Performance on the training trials was above chance in all four age groups: 2.5-year-olds chose the target cup on 47% of trials ($t(23) = 2.853, p = .009$), 3-year-olds on 60% of trials ($t(23) = 4.339, p < .001$), 4-year-olds on 71% of trials ($t(23) = 6.918, p < .001$), and 5-year-olds on 72% of trials ($t(23) = 6.239, p < .001$). Children in all four age groups succeeded on practice trials: they chose the cup that was certain to contain a sticker over two cups that might contain a sticker. This also shows that children at all ages were motivated to get a sticker on their first choice, and were able to follow the two hiding events.
3.3.2. Test Trials

Children virtually never chose the empty cup (this occurred on 3 out of a total 384 test trials, all by 2.5-year-olds). This confirms the results of Experiment 1 and Hill and

![Figure 7](image)

Proportion of training trials (top) and test trials (bottom) in which children in each age group selected the target cup in Experiment 2. Error bars represent 95% confidence intervals, and the dotted line indicates chance (0.33).
colleagues (2012), demonstrating that 2- to 5-year-old children robustly reason by exclusion, even in a more complex version of the cups task. Since children so rarely selected the empty cup, and we were interested in how they would choose among the three remaining options, chance was set at 33%\(^9\). On the three test trials where children picked the empty cup, they were reminded that the cup was empty and asked to choose again; their second choice was treated as their response\(^{10}\).

As in the analysis of training trials, we used the Bonferroni correction for multiple comparisons, resulting in an adjusted alpha of .0125. Three-, 4-, and 5-year-olds chose the target cup significantly more often than chance: 3-year-olds chose the target cup on 58% of trials (\(t(23) = 4.838, p < .001\)), 4-year-olds on 64% of trials (\(t(23) = 4.496, p < .001\)), and 5-year-olds on 76% of trials (\(t(23) = 8.225, p < .001\)). These age groups were more likely to select the target cup than the other two cups, suggesting that they had represented the OR relation between each pair of cups and then reasoned through a disjunctive syllogism. In contrast, 2.5-year-olds chose the target cup on only 36% of trials, which was not different from chance (\(t(23) = 0.425, p = .675\)). They behaved in the manner predicted by the non-

\(^9\) It is possible that the three remaining cups may not have been equally likely choices for children – for example, children might have been more likely to choose the target cup since their attention had just been drawn to that side of the table, or less likely to choose the target cup since that pair of cups had just been “claimed” by their competitor. Given that arguments could be made for both possibilities, we chose to assume a uniform prior. This choice was supported by the finding that 2.5-year-olds’ choices were distributed equally across the three cups, indicating that they were indeed equally likely options.

\(^{10}\) The three empty-cup choices were all by 2.5-year-olds, who were at chance as a group on test trials. Unsurprisingly, this result does not change if these trials are coded as incorrect choices, or if they are dropped from the analysis.
deductive approaches. While they virtually never searched in the empty cup, they chose among the other three cups at chance.

There was no evidence of a learning effect between the first two test trials (57% target cup choices, combining all age groups) and the last two test trials (60% target cup choices) ($t(95) = .818, p = .415$). Trial-by-trial data can be seen in Table 1.

3.3.3. Overall analysis

A two-way ANOVA examined the effects of trial type (training vs. test) and age group (2.5-, 3-, 4-, and 5-year-olds) on percentage of target cup choices. There was a main effect of age ($F(3, 92) = 10.082, p < .001$), but no effect of trial type ($F(1, 92) = 1.420, p = .236$), and no significant interaction ($F(3,92) = .814, p = .489$) between age and trial type. The effect of age on overall performance was confirmed by treating age as a continuous variable: age predicted performance on both training trials ($r(96) = .316, p = .002$) and test trials ($r(96)$

<table>
<thead>
<tr>
<th>Age Group</th>
<th>Trial 1</th>
<th>Trial 2</th>
<th>Trial 3</th>
<th>Trial 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.5-year-olds</td>
<td>33%</td>
<td>38%</td>
<td>30%</td>
<td>42%</td>
</tr>
<tr>
<td></td>
<td>8/24</td>
<td>9/24</td>
<td>7/23</td>
<td>10/24</td>
</tr>
<tr>
<td>3-year-olds</td>
<td>54%</td>
<td>50%</td>
<td>58%</td>
<td>67%</td>
</tr>
<tr>
<td></td>
<td>13/24</td>
<td>11/22</td>
<td>14/24</td>
<td>16/24</td>
</tr>
<tr>
<td>4-year-olds</td>
<td>50%</td>
<td>71%</td>
<td>67%</td>
<td>65%</td>
</tr>
<tr>
<td></td>
<td>12/24</td>
<td>17/24</td>
<td>16/24</td>
<td>15/23</td>
</tr>
<tr>
<td>5-year-olds</td>
<td>67%</td>
<td>86%</td>
<td>75%</td>
<td>78%</td>
</tr>
<tr>
<td></td>
<td>16/24</td>
<td>19/22</td>
<td>18/24</td>
<td>18/23</td>
</tr>
</tbody>
</table>

Table 1

Trial-by-trial data for test trials in Experiment 2, separated by age group. Percentage of target cup choices is shown in grey rows, and raw numbers of target cup choices, out of usable trials, in white rows.
Both types of trials appeared to place heavy demands on the children, leading to lower performance in younger children.

With age as a continuous variable, a partial correlation revealed that age predicted test trial performance, controlling for training trial performance ($r(94) = .374, p < .001$). There was age-related improvement on test trials, over and above that seen on training trials. The gap between test trials and training trials was greater for younger children than older children – this suggests that the test trials were especially difficult for the younger children, while they were no more difficult than the training trials for the older children.

4. General Discussion

Experiment 1 demonstrated that, like individuals from many animal species, 23-month-old toddlers use information about where something is not to constrain a search for where it is. This extends the existing results on reasoning by exclusion in preschoolers and 2-year-olds (Austin et al., 2014; Hill et al., 2012) to children under 2 years of age.

On all three of the interpretations that we considered, reasoning by exclusion requires some way of implementing the thought NOT A. How might this be accomplished? If children are truly reasoning using the disjunctive syllogism, they would represent a logical, abstract symbol NOT that interacts with the logically structured thought “A OR B”. However, if they are not using the disjunctive syllogism, several simpler possibilities remain open. The negation information could be implemented through an operation of “crossing out” or eliminating one possibility. This would not require an explicit symbol for negation that can combine with and enter hierarchically into logically structured thoughts, but rather be a more limited computation of negation. Another possibility is that negation is not
represented at all in these studies – instead, the subject represents the positive thought “A IS EMPTY”. On this account, the representations being used are specific to emptiness, rather than generalizable to other situations involving negation.

Having established that children as young as 23 months old can succeed at the cups task, Experiment 2 asked which mechanism underlies preschoolers’ reasoning by exclusion. Three- to 5-year-old children updated their belief that a sticker was in cup B upon seeing it was not in cup A, indicating that they had represented the disjunctive OR relation between them. These results corroborate the finding that 4- to 6-year-old children use the disjunctive syllogism in an invisible displacement task (Watson et al., 2001) using a novel method, and show that 3-year-olds also have this ability.

Although we were seeking evidence that children were certain that the target cup contained a sticker on test trials, it is also possible that children chose the target cup because it was merely more certain to contain a sticker than the other options. Our design did not allow us to distinguish between a choice based on absolute certainty and one based on increased certainty. The latter would still require that children represented the dependent relationship between the two locations, and that they inferentially updated their assessment of one cup upon seeing that the other was empty; however, the inference children made would not be truly deductive. This possibility was put forth by Rescorla (2009), who described it in a Bayesian framework, where the probability associated with one possibility is adjusted up as the probability of another possibility goes down. However, one feature of our data suggests that children were making a deductive inference: 3- to 5-year-old children chose the target cup just as often in test trials as they did in training trials, in which they could directly observe that a sticker was being hidden there.
The finding that 3- to 5-year-olds could choose an option that necessarily contained a reward over those that might have contained a reward is somewhat surprising in the context of a large literature on children’s difficulty with the concept of logical necessity. However, many of these previous studies asked children to make explicit metacognitive judgments about their certainty – for example, by choosing between “true”, “false”, and “can’t tell” judgments of statements – which required that they distinguish between logical necessity, validity, and truth (e.g. Horbin & Acredolo, 1989; Morris & Sloutsky, 2002; Osherson & Markman, 1975; Russell & Haworth, 1987). In contrast, our study did not depend on making explicit metacognitive judgments and did not require children to understand and assess complex sentences. Our findings are consistent with studies showing that children can compute degrees of belief and monitor certainty – for example, by distinguishing between guessing and knowing – as early as the preschool years (Cultice, Somerville, & Wellman, 1983; Lyons & Ghetti, 2011; Miscione, Marvin, O’Brien, & Greenberg, 1978; Moore, Bryant, & Furrow, 1989; Pillow & Anderson, 2006).

In contrast, the 2.5-year-olds showed no evidence of using the disjunctive syllogism. Instead, their performance appeared to reflect the use of a non-deductive approach that did not implement the concept of OR: they searched equally in all three cups. These data converge with the lack of evidence for using the disjunctive syllogism at this age in the tube-searching task (Call & Carpenter, 2001). Further, they provide an existence proof that young children sometimes fail to represent or reason using the dependent OR relationship between options, and raise the question of whether the toddlers in Experiment 1, as well as the nonhuman animals that have succeeded on the cups task, may also be using a non-deductive strategy rather than making a logical inference. This possibility should be
examined in future research of the developmental and phylogenetic origins of logical reasoning. Importantly, these results underline the need for caution in taking success on the cups task as evidence for representations of logical concepts, or of reasoning by logical inference.

What might explain 2.5-year-olds’ failure to use the disjunctive syllogism in Experiment 2? An examination of children’s performance on the training trials rules out several possibilities. Training trials had the same structure as test trials, and presented children with the same ultimate choice: pick between a cup that is sure to contain a sticker and two cups that may or may not contain stickers. These similarities mean that training trials had many of the same task demands as test trials. The 2.5-year-olds’ success on training trials shows they were motivated to find a sticker on their first choice, could follow and remember two separate hiding events across two different sets of locations, and could choose the sure choice. Since the 2.5-year-olds were successful on training trials – and age predicted success on test trials even controlling for training trial performance – their at-chance performance on the test trials must hinge at least partly on some dissimilarity between the training trials and test trials.

Two broad possibilities remain open. First, 2.5-year-olds may differ from older children in some general aspect of their cognitive abilities that was taxed more by test trials than training trials. On this hypothesis, test trials had greater performance demands than training trials, such that although 2.5-year-olds may be capable of reasoning according to the disjunctive syllogism in other circumstances, they were overwhelmed by the demands of this specific task. Two candidate differences are the working memory requirements of managing representations of four different cups rather than three, or
representing two disjunctions rather than just one. Indeed, older preschoolers’ performance was far from ceiling, confirming that the task placed high demands on young children. It is very likely that changes in executive function contributed to the developmental changes observed. On some accounts of cognitive development during the preschool years (e.g. Andrews & Halford, 2011; Zelazo & Frye, 1998), domain-general developments in working memory and other aspects of executive function enable children to sharply increase their ability to represent hierarchically organized rules and multiple relations during just these years. One caveat to this possibility is that 2.5-year-olds also failed to spontaneously demonstrate deductive reasoning by the disjunctive syllogism in the tubes task (Call & Carpenter, 2001). The tubes task presented children with only two or three options, and thus had considerably lower working memory demands than our task; despite these differences, 2.5-year-olds did not show evidence of making the disjunctive syllogism inference in either case.

The second broad possibility is that differences in logical representation or inference-making abilities play a role in explaining the different performance of 2.5-year-olds and older children. In this case, 2.5-year-olds’ at-chance performance on test trials, despite succeeding on training trials, could be based on their difficulty in representing or reasoning through the disjunctive syllogism. It is possible that the youngest children failed to establish the initial “A or B” premise in working memory, due to a difficulty in representing the logical concept or. There is suggestive support for this proposal in studies of language production: although children productively use the word “and” in language shortly after their second birthday, they do not say the word “or” until about 3 years of age (Bloom, Lahey, Hood, Lifter, & Fiess, 1980; French & Nelson, 1980; Lust & Mervis, 1980;
Morris, 2008). This might hint that the logical relation or is not a part of infants’ conceptual repertoire. This hypothesis, if supported by further research, would be consistent with proposals that language acquisition plays a crucial role in the development of logical capacities.

Reasoning by the disjunctive syllogism depends on abstract, combinatorial thought. It requires implementing the logical concepts or and not, and combining the logically structured thoughts “A or B” and “not A” in a deductive inference. Our results demonstrate that 3- to 5-year-old children are capable of this kind of reasoning, while we failed to find evidence for this capacity in 2.5-year-olds. While the cause of 2.5-year-olds’ failure to reason through the disjunctive syllogism in this study is unknown, their behavior is nonetheless telling: it demonstrates that, in some circumstances, young children reason by exclusion through a non-deductive process that involves avoidance of the eliminated alternative, without updating the remaining alternative. It is therefore also plausible that nonhuman animals that succeed on the cups task may be doing so on the basis of the same non-logical representations and computations. There is as yet no compelling evidence for successful logical reasoning using the disjunctive syllogism in nonhuman animals or children under 3 years of age, leaving open the question of whether they are capable of the same kinds of flexible, abstract, combinatorial thought that adults have.
Paper 3: Young Children Make Logical Inferences in Causal Reasoning

Shilpa Mody & Susan Carey
1. Introduction

The ability to identify and understand causal relationships is essential to making sense of the world around us. Luckily, young children are up to the challenge, demonstrating a set of increasingly sophisticated expectations about how causal events will unfold, ranging from an understanding of simple physical launching events to the role of agents in causal interactions (e.g. Ball, 1973; Kotovsky & Baillargeon, 2000; Leslie & Keeble, 1987; Muentener & Carey, 2010; Saxe, Tenenbaum, & Carey, 2005; Woodward, Phillips, & Spelke, 1993). Yet in addition to specific knowledge about causality, young children’s causal reasoning also reflects aspects of domain-general reasoning – abilities that they also deploy in other contexts.

For example, toddlers engage in probabilistic reasoning about causal outcomes: they preferentially try to activate a toy using an object that has previously activated it 4 out of 6 times, over objects that have activated it 2 out of 6 or 4 out of 12 times (Waismeyer, Meltzoff, & Gopnik, 2015). Similar probabilistic reasoning is exhibited in other areas of infants’ thought, such as when making inferences about samples and populations (e.g. Denison & Xu, 2010; Teglas, Girotto, Gonzalez, & Bonatti, 2007). Another domain-general reasoning ability that plays a part in causal reasoning is the ability to use inductive reasoning to make an overhypothesis: when shown that two pairs of identical objects activate a toy, but that two pairs of different objects do not activate it, toddlers try to activate it with a novel identical pair, apparently making the generalization that the toy would react to any pair of identical objects (Walker & Gopnik, 2014). Similarly, 9-month-olds infants have been found to be able to make overhypotheses outside of the causal
domain: when shown that objects in three boxes were all identical in shape within a box, though different between boxes, infants were surprised when a fourth box contained variably-shaped objects; apparently, they made the generalization that any box would contain identically-shaped objects (Dewar & Xu, 2010).

In this paper, we turn to a third aspect of domain-general reasoning: logical deductive inference. In particular, we ask whether children make the disjunctive syllogism inference to determine the cause of an event. Specifically, when they see that one of two objects causes a toy to activate, but that one of them alone does not, do they conclude that the other object must be able to activate the toy? This question can help shed light on the breadth of young children’s causal reasoning abilities – do they include access to logical representations and inference-making? Conversely, it can also add to our understanding of young children’s logical reasoning abilities – are their logical representations domain-general, able to be applied flexibly to different areas of thought, including causal inference.

Many studies of causal reasoning in toddlers and preschoolers use a “blicket detector” paradigm. In these studies, children are shown that some but not all objects are able to activate a toy, and their inferences are assessed by asking them to evaluate the objects or intervene on the toy themselves. On key trials – variously termed “exclusion”, “indirect screening-off”, or “recovery from overshadowing” – children are shown that placing two blocks on a blicket detector together causes it to activate, but that placing one of those blocks alone does not cause it to activate. When asked to activate the detector, children between 17 months and 3 years old successfully reason by exclusion: they avoid the inert block, instead attempting to activate it with the other block (Mody, Feiman, & Carey, under review; Sobel & Kirkham 2006; Volter, Sentis, & Call, 2016). In contrast, 15-
month-olds perform at chance (Mody, Feiman, & Carey, under review). These findings indicate that around 1½ years of age, children gain the ability to negate or exclude one option, and as a result act on the other option – a pattern of inference that can certainly be glossed as “A or B, not A, therefore B”.

However, there are two alternative ways that a child could successfully reason by exclusion in this task that do not require the full disjunctive syllogism inference. First, the child might not represent that the blocks are potential causes of the detector at all – that is, they may not be representing the initial “A or B” information at all. Instead, they could act solely upon an aversion to the inert block, and end up choosing the other block by default, because it is the only other thing to do. Second, the child could correctly represent the two blocks as potential causes for detector’s activation, but maintain these representations independently: they could separately represent that A might activate the detector, and that B might activate the detector. By failing to represent the “or” dependency between the two options – either one or the other must be true, so information about one option is relevant to the other option – the child would eliminate A from consideration upon seeing that it was inert, but not correspondingly update their assessment of B. Thus, they would act on the second block based on their initial belief that it might activate the detector, rather than the novel, deductive conclusion that it does activate the detector (see Mody & Carey, 2016, for further discussion).

Both of these alternative lines of reasoning are rational: they lead to the sensible behavior of avoiding the inert block and acting on the other one instead. However, neither approach requires that the options be jointly represented with “or”, and neither leads to a deductive, logical conclusion. Crucially, this latter point can be used to distinguish whether
children are actually making the disjunctive syllogism inference when they reason by exclusion, or if they are using one of the two non-deductive strategies instead. Does learning that one block is inert change children's assessment of the other block, making them certain that it activates the blicket detector?

One potential way to measure children's assessment of B is to ask them to categorize each block as either a blicket (active) or not a blicket (inert). When shown exclusion trials in which A and B together activate the blicket detector but that A alone does not, 3- and 4-year-olds state that B is a blicket 100% of the time (Sobel, Tenenbaum, & Gopnik, 2004), suggesting that they have a high degree of certainty that B is a blicket. However, it is possible that this effect is due to a high baseline rate of categorizing blocks as blickets – children may be predisposed to call them blickets when they are actually unsure of their causal status.

Beckers, Vandorpe, Debeys, and De Houwer (2009) addressed this question by showing 3.5-year-old children that a pairs of blocks (A and B) could activate the toy when placed on it together, but that A alone did not. They also showed children that a second pair of blocks (C and D) also activated the toy when placed on it together. They then directly compared children's assessments of B and C, which, indeed, revealed a high baseline rate of calling blocks blickets: children categorized both B and C as blickets nearly 100% of the time. But crucially, they then asked children to activate the toy using either B or C. If children made the disjunctive syllogism inference, they should conclude that B was certain to activate the toy, while they could not be sure of the causal status of C, and should thus preferentially choose B. However, both non-deductive accounts of reasoning by exclusion – simply avoiding the inert block, or excluding the inert block without updating their
assessment of the block that was paired with it – predict that children should have no reason to choose B more often than C. The results were that the children selected B approximately twice as often as C, demonstrating that seeing “not A” raised their assessment of the causal status of B above that of C. This demonstrates that 3.5-year-old children represent the disjunctive relationship “A or B” and make the disjunctive syllogism inference in a causal reasoning context.

Intriguingly, 3.5 years is also the earliest age for which there is clear evidence for reasoning by the disjunctive syllogism in the context of searching for hidden objects. As in causal reasoning, children can exclude one of two potential locations of a hidden object from about 1½ years of age (17-month-olds: Mody, Feiman, & Carey, under review; 20-month-olds: Feiman, Mody, Sanborn, & Carey, under review; 23-month-olds: Mody & Carey, 2016; 3-, 4-, and 5-year-olds: Hill, Collier-Baker & Suddendorf, 2012). But again, children’s successful reasoning by exclusion is ambiguous – it does not definitively show that they have concluded “therefore B”, and thus that they have made the disjunctive syllogism inference. To investigate this, we showed 2.5- to 5-year-old children that a sticker was hidden in either cup A or cup B, that an identical sticker was hidden in either cup C or cup D, and then that A was empty (Mody & Carey, 2016). We found that children between 2.5 and 3 years of age reliably avoided selecting the cup they had seen was empty, but chose randomly among the other three cups, showing no signs of making the disjunctive syllogism inference. In contrast, 3.5-year-olds, 4.5-year-olds, and 5.5-year-olds preferentially chose the cup that was paired with the empty cup, indicating that seeing “not A” led them to the deductive conclusion “therefore B”. Both the success of older children and the failure of younger children persist in studies using other methods: 2.5-year-olds
fail to spontaneously make the disjunctive syllogism inference to initiate a search, instead continuing to gather information (Call & Carpenter, 2001), while 4- to 6-year-olds search at an increased speed when they can use the disjunctive syllogism to infer that they are searching in the correct location (Watson et al., 2001). Together, these findings suggest that it is only by 3.5 years of age that children are able to use the disjunctive syllogism to infer the location of a hidden object.

In the current study, we ask when the ability to apply the disjunctive syllogism to infer the cause of an event emerges in development. Given the precociousness that infants display in a wide array of causal reasoning tasks, and in particular their ability to reason by exclusion in blicket detector paradigms as early as 17 months, it is possible that very young children bring simple deductive reasoning abilities to the arena of causal reasoning. This would be the earliest evidence for the ability to make deductive inferences in childhood. However, another possibility is that children's ability to make the disjunctive syllogism inference in a causal context mirrors their performance on search tasks, where there is no clear evidence that children use negative information to come to a deductive conclusion until after 3 years of age. If so, this might suggest that the common logical resources that are called on by the two tasks are not simply not available, or are not easily deployed, by children younger than 3 years old.

2. Methods

2.1. Participants

We tested three groups of children: 2.5-year-olds (N = 48, mean age = 30.00 months, range = 27.11-32.96, 25 girls), 3-year-olds (N = 48, mean age = 35.97 months, range =
33.36-38.85, 24 girls), and 3.5-year-olds (N = 48, mean age = 42.17 months, range = 39.14-45.69, 23 girls). Participants were recruited by phone and email from the Greater Boston Area and were tested at the Harvard Lab for Developmental Studies. Families were compensated with a $5 travel reimbursement and a small toy.

A further 23 children were tested but excluded from the final sample for completing fewer than eight out of 11 trials with usable data (18), experimenter error (1), or making no response on the first three trials, which led us to abort testing (4).

2.2. Materials

We constructed a blicket detector using a small cardboard box and a commercially available “magic wand” toy; when the toy was activated, multicolored LEDs would light up and spin around. The toy was attached to the front of the box with the lights facing the child, and could be surreptitiously turned on and off by the experimenter’s hand inside the box.

The other materials were a set of wooden blocks of various shapes and colors. Each trial used two to four same-shape, different-colored blocks; the shape of the blocks, as well as the identity of the target block, varied across trials. The blicket detector and blocks were placed on a tray that could be slid in and out of children’s reach.

2.3. Procedure

Children sat on their caregivers’ lap, across a table from the experimenter. Caregivers were instructed to remain silent during the trials. To begin the study, children were introduced to the blicket detector: they were told that it could be used to find blickets,
that some of the blocks were blickets and would make it go, and that other blocks were not blickets and would not make it go.

On each trial, the experimenter placed that trial’s blocks in a row in front of the bicket 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 one’s a bicket!”, and when demonstrating inert blocks she did not activate it and said “No, this one’s not a bicket”. When demonstrating two blocks simultaneously, she activated the toy and said “Wow, one of these is a bicket!”. Following the demonstration, she returned the blocks to their original positions, then slid the tray holding the bicket detector and blocks towards the child, saying, “Your turn! Can you use a bicket to make it go?”.

We coded children’s responses by which block they placed on the bicket detector first. Children received feedback on their responses: if the child chose a block that was sure to be a bicket (hereafter, the “target block”), the toy activated and the experimenter gave them a sticker. If they chose any other block, the toy did not activate – in all cases, it was plausible that this was not a bicket. If they placed multiple blocks on the detector simultaneously, the toy activated and they received a sticker, but they were told to try only one block on the next trial.

Each child then participated in three warm-up trials, four baseline trials, and four test trials; trials were always presented in this order. The color and position of the active blocks varied across trials and were counterbalanced across children.
2.3.1. *Warm-up trials*

The three warm-up trials were designed to scaffold children’s understanding of the task. Each warm-up trial used two blocks. On these trials only, if children did not respond correctly, they were encouraged to make another response; only their first choice was used in analyses.

On the first warm-up trial, the experimenter demonstrated the target block, which activated the toy, and explained “Wow, this makes it go! That’s because this one’s ablicket”. Next, she demonstrated the other block, which did not activate the toy, and explained “No, this doesn’t make it go. That’s because this one is not ablicket”. She then asked the child to activate the toy. At the end of the trial, she explained that she would be giving them a sticker each time they activated the toy for the rest of the study.

The second warm-up trial was identical to the first, except that when demonstrating the blocks, the experimenter simply said “Wow, this makes it go!” or “No, this doesn’t make it go”, then asked the child whether they thought each block was ablicket. Regardless of what children said, she told them the correct answer. She then asked the child to activate the toy.

The third warm-up trial introduced children to the effect of placing two blocks on the toy simultaneously. The experimenter demonstrated the target block, which activated the toy, then demonstrated the other block, which did not activate the toy. She then placed both blocks on theblicket detector simultaneously, and explained “Wow, this makes it go! That’s because one of these is ablicket”. She then asked the child to activate the toy.
2.3.2. *Baseline trials*

The four baseline trials required that children understand that a demonstration of two blocks placed on the blicket detector simultaneously meant that they could not know which of the blocks was a blicket. They were designed to measure children’s tendency to select a block that was sure to be a blicket (the target block) over blocks that might be blickets. This was an essential component of successful performance on test trials.

Each trial used three blocks, arranged such that one pair of blocks was on one side of the tray and the target block was on the other side of the tray. On two of the baseline trials, the experimenter first placed the target block on the detector, which activated it, then placed the pair of blocks on the detector, which also activated it; these two events were reversed for the other two baseline trials. She then asked the child to activate the toy.

2.3.3. *Test trials*

The four test trials asked children to reason using the disjunctive syllogism. If children made this inference, they would be faced with essentially the same choice as in baseline trials: one block that was sure to be a blicket (the target block) and two blocks that might be blickets, as well as one block that was not a blicket.

We created two versions of the test trials, *negation-middle* and *negation-last*, which varied when in the trial the negative information was presented; this was manipulated between subjects. Each trial used four blocks, arranged with one pair of blocks on each side of the tray.

In the negation-final condition, the experimenter placed one pair of blocks on the detector, which activated it, then placed the other pair of blocks on the detector, which also
activated it. Finally, she placed one block on the detector by itself (the *negative block*), which did not activate it. The negative block was from the first-demonstrated pair on two of the test trials, and from the last-demonstrated pair on the other two test trials. She then asked the child to activate the toy. In the negation-middle condition, the experimenter placed one pair of blocks on the detector, which activated it, then placed the negative block on the detector, which did not activate it. The negative block was always one of the blocks in the first pair. Finally, she placed the other pair of blocks on the detector, which activated it.

We constructed these two versions because we anticipated that children might be predisposed to select a block from the last pair that the experimenter manipulated. Children’s attention had just been drawn to this pair, so they might choose a block from this pair simply because their attention lingered there. Further, on negation-middle trials, the experimenter gave them the information required to find a certain blicket, but then demonstrated the other pair of blocks just prior to asking them to activate the detector; these pragmatic cues could lead children to believe that the experimenter was asking them to find the blicket among this second pair of blocks. Conversely, on negation-final trials, the experimenter demonstrated the negative block just before asking children to find the blicket, which could lead children to believe that the experimenter was asking them to find the blicket among this pair of blocks. If attentional or pragmatic cues influenced children’s choices, it would lead children to be biased *towards* the target block on negation-final test trials, and biased *away from* the target block on negation-middle test trials. We thus tested half the children in each condition, and averaged the results.

If children made the disjunctive syllogism inference, they should infer that the target
block, which was paired with the negative block, must be a blicket. They should therefore preferentially try to activate the blicket detector with the target block, since it is more likely to activate it than the two blocks that may or may not be blickets. However, if children excluded the negative block without making any inference about the block paired with it, they should demonstrate no preference for the target block compared to the other two blocks.

3. Results

For 13 of the 144 children, one trial was excluded due to the child's unwillingness to respond (12) or experimenter error (1). A further 79 trials – spread across 19 2.5-year-olds, 20 3-year-olds, and 12 3.5-year-olds – were excluded due to the child placing two or more blocks onto the detector simultaneously. While this is a rational strategy, increasing the probability of activating the detector, it is difficult to interpret.11

Using the remaining trials, the percentages of target block choices were computed for each child (see Figure 8). For all comparisons of the individual age groups (i.e. one-sample t-tests, binomial tests, simple effects), we used the Holm-Bonferroni correction for

11 Including the 79 trials on which children placed two or more blocks on the blicket detector simultaneously (coded as incorrect choices) had essentially no effect on the overall pattern of results. Children in all age groups succeeded on baseline trials (2.5-year-olds: t(47) = 7.418, p < .001; 3-year-olds: t(47) = 6.310, p < .001; 3.5-year-olds: t(47) = 5.645, p < .001). However, on test trials, 2.5-year-olds’ performance was indistinguishable from chance (t(47) = 0.252, p = .803), 3-year-olds were marginally better than chance (t(47) = 2.313, p = .050), and 3.5-year-olds were significantly better than chance (t(47) = 4.756, p < .001). The pattern of significant effects and interactions in the three-way ANOVA were also unchanged from those reported in the main text.
multiple comparisons, and present adjusted p-values below.

3.1. Warm-up trials

We established chance at 50%, since there were two blocks to choose from. Children in all three age groups performed nearly at ceiling on warm-up trials: 2.5-year-olds chose the target block on 96.5% of trials ($t(47) = 31.326, p < .001$), 3-year-olds on 94.1% of trials ($t(47) = 20.129, p < .001$), and 3.5-year-olds on 97.2% of trials ($t(47) = 35.140, p < .001$). This demonstrates that children were highly motivated to activate the detector and
understood the causal demonstrations.

### 3.2. Baseline trials

We established chance at 33%, as there were three blocks to choose among. Performance on the baseline trials was above chance in all three age groups: 2.5-year-olds chose the target block on 68.3% of trials (t(47) = 7.967, p < .001), 3-year-olds on 61.5% of trials (t(47) = 6.719, p < .001), and 3.5-year-olds on 58.7% of trials (t(47) = 6.195, p < .001). Furthermore, children did well on their very first baseline trial: the target block was chosen by 28 out of 45 2.5-year-olds\(^{12}\) (one-tailed binomial test based on 1/3 chance rate, p < .001), 20 out of 43 3-year-olds (p = .050), and 24 out of 44 3.5-year-olds (p = .006). Combining all age groups together, there was no evidence of a learning effect between the first two baseline trials (60.1% target block choices) and the last two baseline trials (65.6%) (t(141) = 1.315, p = .191). Children in all three age groups succeeded on baseline trials: they chose the block that was sure to be a blicket over the two blocks that might have been blickets.

### 3.3. Test trials

Since we were primarily interested in whether children would choose the target block more often than the two blocks that might have been blickets, we established chance at 33%. This was a conservative choice, as children – particularly the youngest ones – did sometimes select the block we had shown them was inert: 2.5-year-olds did so on 37 trials, 3-year-olds on 18 trials, and 3.5-year-olds on 11 trials.

\(^{12}\) This and all further analyses of first trial performance exclude children who did not generate usable data on the trial.
3-year-olds on 21 trials, and 3.5-year-olds on 11 trials.

We found that 3- and 3.5-year-olds succeeded on test trials: 3-year-olds chose the target block on 45.8% of trials ($t(47) = 3.167, p = .006$) and 3.5-year-olds chose it on 53.8% of trials ($t(47) = 4.961, p < .001$). Furthermore, these children did well at their very first test trial: the target block was chosen by 23 out of 43 3-year-olds (binomial test based on 1/3 chance rate, $p = .010$), and 26 out of 47 3.5-year-olds ($p = .006$). Children at these ages were more likely to select the target block than the blocks that might have been blickets, suggesting that they had used the information about the inert block to reason through a disjunctive syllogism.

However, 2.5-year-olds chose the target block on only 33.3% of trials, which was identical to chance ($t(47) = 0.000, p = 1.000$). They performed at chance on their first test trial, with 17 out of 44 children selecting the target block ($p = .275$). There was no indication that 2.5-year-olds were reasoning through the disjunctive syllogism: they demonstrated no preference for the target block.

One possible reason for 2.5-year-olds’ poor performance on test is that they chose the negative block at a high rate, which we coded as an incorrect response; yet choosing this block indicates that they did not even reason by exclusion. We thus reran this analysis, including only the trials on which children did not select the negative block: 2.5-year-olds chose the target block on 42.9% of trials, which was marginally better than chance ($t(45) = 1.946, p = .058$). This weak success leaves open the possibility that when 2.5-year-olds

---

$^{13}$ Two 2.5-year-olds generated no usable test trials, and were dropped from this analysis.
successfully exclude an inert block, they may sometimes, though not reliably, use the disjunctive syllogism to do so.

Again, combining all age groups together, there was no evidence of a learning effect between the first two test trials (45.0% target block choices) and the last two test trials (44.0%) \( t(140) = 0.243, p = .808 \). While 3- and 3.5-year-olds showed evidence of making the disjunctive syllogism inference on test trials, 2.5-year-olds did not.

### 3.4. Overall analysis

We conducted a two-way mixed-measures ANOVA to examine the effects of trial type (baseline/test) and age group (2.5-/3-/3.5-year-olds). As hypothesized, there was a main effect of trial type \( F(1, 141) = 33.322, p < .001 \), with children performing better on baseline trials than test trials. There was no main effect of age. We also found an interaction between trial type and age group \( F(2, 141) = 7.547, p = .001 \) – children’s performance on test trials (which required a disjunctive syllogism inference) versus baseline trials (which did not) varied by age. Analyzing simple effects, we found that 2.5-year-olds performed better on baseline trials than test trials \( F(1, 141) = 39.69, p < .001 \), as did 3-year-olds \( F(1, 141) = 7.96, p = .010 \). However, 3.5-year-olds performed equivalently on the two types of trials \( F(1, 141) = 0.77, p = .382 \) – they paid no cost for making the disjunctive syllogism inference.

### 3.5. Effect of negation timing on test trials

We conducted a two-way ANOVA examining the effect of negation timing (negation-middle/negation-final) and age (2.5-/3/3.5-year-olds) on children’s performance on test
trials. We found that performance increased with age (F(2, 138) = 7.168, p = .001). We also found a main effect of negation timing (F(1, 138) = 12.494, p = .001), due to better performance on negation-final test trials than negation-middle test trials. Finally, we found an interaction between age group and negation timing (F(2, 138) = 3.829, p = .024); analyzing simple effects, we found that 2.5-year-olds did significantly better on negation-final than negation-middle test trials (F(1, 138) = 12.329, p = .003), as did 3-year-olds (F(1, 138) = 7.790, p = .012); in contrast, 3.5-year-olds performed equivalently on the two kinds of test trials (F(1, 138) = 0.032, p = .857). To further unpack this finding, we analyzed the children who had participated in negation-middle test trials versus negation-final test trials separately.

3.5.1. **Negation-middle test trials**

We first considered children who had participated in negation-middle test trials (see Figure 9, panel A). On these test trials, children’s attention had last been drawn to the incorrect pair of blocks, so we expected their performance on test trials might be biased away from the target block.

Indeed, when we compared children’s performance on test trials to chance, we found that 2.5-year-olds performed below chance – they selected the target block on only 19.8% of trials (t(23) = 4.175, p < .001). Three-year-olds chose the target block on 35.1% of trials, which was indistinguishable from chance (t(23) = 0.405, p = .689), while 3.5-year-olds chose the target block on 54.5% of trials, which was better than chance (t(23) = 3.662, p = .002). When the experimenter drew children’s attention to the incorrect pair of blocks last, only the oldest children showed a preference for the target block, apparently
overcoming the attentional or pragmatic lure of the incorrect blocks.

We then examined the overall pattern of data in this subset of children, using a two-way mixed-measures ANOVA. This revealed a main effect of trial type ($F(1, 69) = 44.926, p < .001$), with children performing better on baseline than test trials, and a main effect of age ($F(2, 69) = 3.394, p = .039$). As in the full sample, there was an interaction between age group and trial type ($F(2, 69) = 9.372, p < .001$). Analyzing simple effects, we again found that 2.5-year-olds performed better on baseline trials than test trials ($F(1, 69) = 53.17, p < .001$), as did 3-year-olds ($F(1, 69) = 8.56, p = .010$). In contrast, 3.5-year-olds performed just as well on test trials as baseline trials ($F(1, 69) = 1.94, p = .169$).

Figure 9
Percentage of baseline and test trials on which children in each age group tried to activate the blicket detector with the target block, grouped by whether children participated in negation-middle test trials (panel A, left) or negation-final test trials (panel B, right). Error bars represent 95% confidence intervals, and the dotted line indicates chance (33%).
3.5.2. *Negation-final test trials*

We conducted the same analyses with the children who had participated in negation-final test trials (see Figure 9, panel B). On these test trials, children’s attention had last been drawn to the negative block, which was paired with the target block – we thus expected that children’s performance might be biased *towards* the target block.

When we compared children’s performance on test trials to chance, we found that 2.5-year-olds’ chose the target block on 46.9% of trials, which was marginally better than chance (*t*(23) = 2.025, *p* = .055). Three-year-olds chose the target block on 56.6% of trials, which was better than chance (*t*(23) = 3.920, *p* = .003), and 3.5-year-olds chose it on 53.1% of trials, which was also better than chance (*t*(23) = 3.290, *p* = .006). When the experimenter last drew children’s attention to the negative block, which was paired with the target block, children at all ages showed a preference for the target block.

When we examined the overall pattern of data in this subset of children using a two-way mixed-measures ANOVA, we found only a marginal main effect of trial type (*F*(1, 69) = 3.839, *p* = .054), with children performing better on baseline trials than test trials. There was no significant effect of age, and no interaction.

These results suggest that two factors impacted children’s performance on test trials. First, children became better at making the disjunctive syllogism inference with age: 3.5-year-olds chose the target block on test trials just as often as they did on baseline trials, 3-year-olds performed above chance on test trials but not as well as they did on baseline trials, while 2.5-year-olds chose the target block at chance levels on test trials despite succeeding on baseline trials. Second, children became more resistant to attentional/pragmatic cues that could bias their choices towards the pair of blocks that the
experimenter manipulated last: 3.5-year-olds’ performance was identical on negation-middle and negation-final test trials, while both younger age groups performed better on negation-final test trials, in which the target block was in the last-manipulated pair. This effect was particularly pronounced for the 2.5-year-olds, who actually performed below chance on negation-middle test trials; this suggests that their apparent success on negation-final test trials was due entirely to attentional/pragmatic bias.

4. Discussion

In this study, we explored when children begin to make logical, deductive inferences to determine the cause of an event. We showed children that one of each of two pairs of blocks (A or B, C or D) activated a blicket detector, but that one specific block did not activate it (not A). When asked to use one of the blocks to activate the detector, 3- and 3.5-year-old children preferentially chose the target block – which had been paired with the negative block – indicating that they combined the information they had seen to conclude that this block must be able to activate the toy (therefore B).

Children at these ages succeeded on their first test trial, demonstrating that their success was not driven by feedback or learning over the course of the study. The performance of the 3.5-year-olds was particularly robust: when we analyzed children who participated in negation-final and negation-middle conditions separately, each group performed better than chance, which provides an internal replication. Furthermore, 3.5-year-olds’ performance on test trials was identical to their performance on baseline trials, indicating that these children paid no cost for making the disjunctive syllogism inference. While 3-year-olds’ performance was somewhat less strong, they preferentially chose the
target block when shown negation-final test trials, as well as when all children’s data was
grouped together; they performed at chance levels only when shown test trials in which
their attention was last drawn towards the incorrect pair of blocks, a situation that we
predicted would bias the results away from a correct choice. Our results thus corroborate
the findings of Beckers and colleagues (2009), and extend them to children six months
younger: by 3 years of age, children make the disjunctive syllogism inference to deduce the
cause of an event, suggesting that domain-general logical resources are available for causal
reasoning.

One caveat to this conclusion is that the inference that 3- and 3.5-year-olds made
might not have been truly deductive, but could instead be probabilistic. These children’s
preferential choice of the target block might indicate that they were more certain that it
was a blicket than the other blocks, not necessarily that they were completely certain. A
probabilistic inference would still require that children represent the disjunctive
dependency between the blocks, and that they update that their assessment of the target
block based on the information that the block paired with it was inert, but would not lead
to deductive certainty. Many researchers have argued that causal reasoning as a whole
reflects just such probabilistic processes (e.g. Gopnik & Wellman, 2012; Tenenbaum, Kemp,
Griffiths, & Goodman, 2011).

One might take the relatively low rate of success on test trials as evidence that
children are in fact reasoning probabilistically – if they were truly certain that the target
block was a blicket, they should choose it on every trial. However, children’s performance
on baseline trials suggests that this is not the case. On these trials, they were given
unambiguous evidence that the target block was a blicket: it activated the toy, and the
experimenter said “Wow, this one’s a blicket!”. Yet children only selected the target block on roughly 60% of baseline trials; apparently, this represents the ceiling on children’s performance. And 3.5-year-olds’ performance on test trials was indistinguishable from their performance on baseline trials, indicating that making the disjunctive syllogism inference resulted in hitting this ceiling. It is likely that the metacognitive demands of choosing a sure choice over an uncertain one, or distinguishing between guessing and knowing, contributed to the overall low performance (e.g. Lyons & Ghetti, 2011; Morris & Sloutsky, 2002; Pillow & Anderson, 2006). Children may also have been motivated to explore the blocks whose causal status was unknown, rather than simply to activate it through known means (Gweon & Schulz, 2008; Schulz & Bonawitz, 2007).

In contrast to the older children, 2.5-year-olds performed at chance on test trials, choosing the target block no more often than the two blocks that might have been blickets. While 2.5-year-olds who participated in negation-final test trials were marginally successful, those who participated in negation-middle test trials actually performed worse than chance. The attentional and pragmatic cues provided by the experimenter towards either the correct or incorrect pair of blocks had a large effect on these children’s performance, suggesting that any hint of success found in the negation-final condition was driven by these cues. At-chance performance on test trials is predicted by both non-deductive approaches to reasoning by exclusion: these children may have represented the possibility that each block was a blicket independently, and thus failed to update their assessment of the target block upon learning that the block paired with it was inert, or they may have simply avoided the negative block. Either of these approaches would lead to successful reasoning by exclusion, as expected based on previous research (Mody, Feiman,
& Carey, under review; Sobel & Kirkham, 2006; Volter, Sentis, & Call, 2016), but neither
involves reasoning through the disjunctive syllogism. Our results thus provide no evidence
that 2.5-year-olds’ make logical, deductive inferences in the context of causal reasoning.

What might have led to 2.5-year-olds’ chance performance on test trials? An
examination of their performance on baseline trials rules out some possibilities. Baseline
trials were very similar to test trials – for example, they required that children choose a
block that was certainly a blicket over blocks that might be blickets, that they understood
the disjunctive information the experimenter conveyed by placing two blocks on the
detector simultaneously while saying, “Wow, one of these is a blicket!”, and that they were
motivated to activate the detector – yet 2.5-year-olds succeeded, performing significantly
better on baseline trials than test trials. Thus, while these factors may have led to an overall
dampening of performance on both kinds of trials, some additional feature – specific to test
trials and not shared with baseline trials – is required to explain 2.5-year-olds’ particularly
poor performance on test trials.

One possibility is that children at this age are able to make the disjunctive syllogism
inference in causal reasoning situations, but additional factors related to the complexity of
test trials overwhelmed or confused them, leading them to revert to a simpler non-
deductive line of reasoning. For example, the youngest children’s working memory may
have been taxed by the requirement to represent the causal status of four individual blocks,
or two disjunctive relationships. One feature of the data supports this proposal: 2.5-year-
olds selected the negative block on fully 19% of test trials, indicating that they were not
even reasoning by exclusion on these trials, perhaps due to being overwhelmed. When
these trials were removed from the analysis, 2.5-year-olds’ performance improved to be
marginally better than chance, hinting that they may make the disjunctive syllogism inference in some cases.

This hint of success and the high rate of selecting the negative block stand in contrast to our prior study of searching, in which children between 2.5 and 3 years of age selected the cup they had been shown was empty on only 3% of test trials, and performed no better than chance regardless of whether those trials were included in the analysis (Mody & Carey, 2016). In the case of search, the lack of evidence for reasoning by the disjunctive syllogism below age 3 was also found in a study that had much lower working memory demands (Call & Carpenter, 2001), making it less likely that task factors could fully explain those children’s performance. Analogously, the possibility that children younger than 3 years old are able to use the disjunctive syllogism in causal reasoning could be further pursued using tasks with reduced task demands; for example, by comparing children’s preference for a block that is deductively certain, by the disjunctive syllogism, to activate a blicket detector to a single block that activates it 50% of the time.

A second, non-mutually-exclusive possibility is that children younger than 3 years old performed at chance on test trials because they are unable to apply domain-general logical resources to the context of causal reasoning. Strikingly, this is also the earliest age for which there is evidence that children make that same inference in search (Call & Carpenter, 2001; Mody & Carey, 2016; Watson et al., 2001). This raises the possibility that the reason that younger children do not apply logical faculties in causal situations is that they simply do not have the ability to use logical representations to make deductive inferences until age 3. The parallel developmental pattern in the domain of search suggests that our findings may reflect not only which logical resources are available to young
children’s causal reasoning, but which logical resources are available to them at all.
Conclusion
1. Summary of findings

In Paper 1, we explored infants’ access to a flexible, domain-general concept of negation by examining their ability to reason by exclusion across different contexts. In Experiments 1 and 2, a toy was hidden in one of two buckets, then one bucket was revealed to be empty. When asked to find the toy, 17-month-olds, but not 15-month-olds, used the negative information and approached the other bucket. In Experiment 3, two blocks placed on a toy together activated it, then one of those blocks was shown to not activate the toy. Again, 17-month-olds, but not 15-month-olds, used the negative information and attempted to activate the toy with the other block. We argued that the parallel emergence of the ability to reason by exclusion for searching and causal reasoning – and, based on preexisting evidence, word-learning – suggests that infants’ performance hinges on a common resource, likely their ability to flexibly deploy negation in reasoning.

In Papers 2 and 3, we looked for evidence that children’s performance on these tasks actually reflected a deductive inference, rather than a simpler non-deductive strategy that did not require representing the disjunctive relation between the options. In Paper 2, we found evidence that 3- to 5-year-olds used the disjunctive syllogism in a search task: when they saw that a sticker was hidden in one of two cups, and that one cup was empty, their choices appeared to reflect deductive certainty that the other cup contained the sticker. In contrast, 2.5-year-olds did not appear to have used the negative information (that one cup was empty) to update their assessment of the other cup. In Paper 3, we found similar results in a causal reasoning task: when 3- and 3.5-year-old children were shown that one of two blocks caused a toy to activate, but that one of those blocks did not activate
the toy by itself, their choices appeared to reflect deductive certainty that the other block could activate the toy. Again, 2.5-year-olds did not appear to change their assessment based on negative information. These results demonstrate that children use the disjunctive syllogism in both searching and causal reasoning contexts by the age of 3, but that younger children may not easily deploy this logical inference.

2. **The disjunctive syllogism and Bayesian probabilistic inference**

One account of reasoning by exclusion that was not addressed in the introduction is the possibility that infants and animals could succeed by reasoning probabilistically, rather than deductively (Rescorla, 2009). For example, Rescorla (2009) describes a hypothetical situation where a dog is chasing after its prey, and reaches a three-way fork in the road. After sniffing the first two paths, and smelling nothing, it speeds down the third path without pausing to sniff it first. However, rather than ascribing deductive reasoning to this hypothetical dog, Rescorla argues that it could represent the possible locations of its prey in three iconic, map-like representations of the crossroads, with its prey located down a different fork in each map. These maps could then form a hypothesis space over which the dog could engage in probabilistic, Bayesian reasoning. When it sniffs down a path and smells nothing, the probability associated with the relevant map lowers, and the probabilities of the remaining maps increase. After two sniffs, the probability associated with the third map is high enough to clear the threshold for action, and the dog starts down that path.

Probabilistic Bayesian inference provides a parsimonious explanation for infants’ exclusion performance, in the context of a wide literature suggesting that many aspects of
early thought can be modeled as Bayesian inference (e.g. Frank & Tenenbaum, 2011; Gopnik & Wellman, 2012; Teglas et al., 2011; Tenenbaum, Kemp, Griffiths, & Goodman, 2006). Yet normative reasoning, in at least some cases, is better captured by the all-or-nothing character of deductive inference – after all, in the search task, the ball really is certain to be in the second bucket, not merely likely to be.

An open question is whether Bayesian reasoning is truly an alternative to reasoning by the disjunctive syllogism, or one way of implementing it. The construction of the hypothesis space requires that children enumerate the relevant possibilities, and the inference mechanics maintain a fundamentally disjunctive relation between them. Further, the lowering of probability associated with gaining negative information essentially implements negation. Thus even if reasoning proceeds probabilistically, propositional representations including negation and disjunction might be required to represent the information that the probabilistic mechanism uses as input.

3. More on negation

Across referent disambiguation, search, and causal reasoning, infants begin to successfully reason by exclusion at around 1.5 years of age, suggesting that this reasoning recruits a domain-general representation of negation. Yet it is not until age 3 that they show evidence of making the full, deductive disjunctive syllogism inference. So when 17-month-olds successfully reason by exclusion, what might their concept of negation be like?

One possibility sketched out in Papers 1 and 2 is that infants may be recruiting a procedure that implements some limited uses of negation, rather than a freely combinable symbolic representation. This “option-elimination” procedure could operate over
representations of possible states of affairs, without those representations necessarily
having to be propositional – for example, an imagistic working memory model of the toy
hidden in bucket A. This procedure could work across domains – as long as the possible
options could be represented, the procedure would be agnostic to the content of those
representations, needing only to be able to eliminate one of them from consideration. Given
that children do not appear to understand the truth-functional meaning of the words “no”
and “not” until 24 to 27 months of age, it is plausible that they might not have access to the
symbolic representation of negation that these words presumably map on to (Austin et al.,
2014; Feiman et al., under review).

So how might we tell if infants really are using an option-elimination procedure to
reason by exclusion, versus an adult-like symbol for negation? A symbolic negation
representation would be much more flexible; in particular, it could be embedded inside
other functions. Thus, we might see if 17-month-old infants are able to use negation in
situations where it is not used for reasoning by exclusion, and is instead input to another
function.

For example, we recently conducted a study on 14-month-olds’ ability to perform
the classic match- and nonmatch-to-sample tasks (MTS and NMTS, Hochmann, Mody, &
Carey, 2016). Infants saw three identical cards arranged in a row, which flipped over to
reveal symbols: the card in the middle was the same as one of the side cards, and different
from the other side card. In the MTS condition, infants were shown that the side card that
was the same as the middle card would animate, while in the NMTS condition, the side card
that was different from the middle card would animate. We found that children could learn
to predict which card to look at in both MTS and NMTS.
However, rather than using a representation of “same” for MTS and a representation of “different” or “not-same” for NMTS, they seemed to rely only on representations of “same” for both cases – they learned to “look at same” in MTS, and “avoid same” in NMTS. If 17-month-old infants have access to a flexible, combinable concept of negation, they might be able to deploy it in NMTS: instead of only learning that they should “avoid same”, they might also learn to “look at not-same”.

4. More on disjunction

In the section above, I sketched a possible account for the disparity between the emergence of reasoning by exclusion at 1.5 years of age, and the use of the disjunctive syllogism at 3: the representation of negation used by very young children might not be symbolic and easily combined with other representations. But why else might children fail to show evidence of making the disjunctive syllogism inference 3 years of age? One possible reason, which we argue for in Paper 2, is that until this age, children may not have ready access to the logical concept of disjunction. If children are not able to represent the disjunctive relation between possibilities, they would have no way of appropriately updating their assessment of B upon finding “not A”.

However, in an ongoing study (with Sophia Clavel and Susan Carey), we are finding evidence that children may understand the meaning of the word “or” as early as 2 or 2.5 years old. In this study, children are shown three different familiar objects on each trial, and are asked to give some of them to a stuffed bear with a request that includes either the word “or” or “and”. For example, they might hear, “Can you give Mr. Bear the truck or the ball?” or “Can you give Mr. Bear the truck and the ball?”. The results show that for 2.5- and
3-year-olds, the dominant response to “or” questions is to give one of the named objects, and the dominant response to “and” questions is to give both of the named objects. While 2-year-olds made numerous errors, such as giving all the objects on the table, children at all three ages clearly differentiated between the requests, rarely making “and” responses (giving both items named) when asked an “or” question, or vice versa.

If children understand the meaning of “or” – as demonstrated by knowing that giving just one of the objects fulfills a request to “give X or Y” – they clearly have some grasp of the concept of disjunction. Further studies are required to more fully clarify the extent of this knowledge, particularly in light of recent findings suggesting that much older children may still have difficulty resolving simple sentences containing “or” (Singh, Wexler, Astle, Kamawar, & Fox, 2015; Tieu et al., 2016). But if we assume that 2.5-year-olds really can represent disjunction, what else might lead to their failure to reason through the disjunctive syllogism?

One alternative is that very young children can represent the disjunctive relation between two possibilities, but it does not naturally occur to them to do so. This would predict that children might succeed on the tasks in Paper 2 and 3 if they were simply told how to represent the relation: “The sticker is in the red cup or the blue cup!” or “Wow, either the red block or the blue block makes it go!”.

A second alternative is that children can represent the disjunctive relation between two possibilities, but that children face difficulties when trying to combine this information with negation. Both a disjunctive structure (A or B) and a negative structure (not A) require that the logical concept be combined with a representation of a state of affairs. But combining these two structured representations in inference further requires the ability to
form a hierarchically structured representation, where two combined constituents are themselves combined into a larger representation. This resembles the idea of “relations between relations” that was suggested as the source of discontinuity between animal and human thought by Penn, Holyoak, and Povinelli (2008).

A third alternative is that 2.5-year-olds may be able to make the disjunctive syllogism inference, yet may have been blocked from doing so by other demands of our tasks. The four-cup and four-block designs used in Papers 2 and 3 likely placed high demands on children’s working memory and metacognition, as evidenced by the overall low rate of success. Ideally, we would find ways to measure children’s certainty in the conclusion reached by reasoning exclusion, using simplified paradigms.

Tentatively, one potential avenue is to see how certain children are about the referent of a word they are introduced to via referent disambiguation: does ruling out the car as the potential referent to “dax” lead children to be sure that the speaker was referring to the novel object? In particular, does this lead them to be just as sure as they would be if the novel object was presented alone, and thus was the only available referent for “dax”? This certainty could be assessed by seeing how well children remember the mappings between novel labels and novel objects.

Prior studies have shown that although children increase their looking to the novel object on referent disambiguation trials by 18 months of age, and are able to remember novel object-label pairings when heard in isolation, they do not show evidence of remembering those novel object-label pairings at either 18 or 24 months (Bion, Borovsky, & Fernald, 2012; Horst & Samuelson, 2008). By 30 months of age, children show signs of remembering mappings learned by referent disambiguation, indicating that they at least
considered the novel object as a possible referent for the novel word; but crucially, these studies have not assessed whether that retention reflects the same level of certainty as if the novel object had been the only one present (Bion et al., 2012; Golinkoff, Hirsh-Pasek, Bailey, & Wenger, 1992; Horst, Scott, & Pollard, 2010; Spiegel & Halberda, 2011). In contrast, at 38 months of age, children’s retention of novel object-label pairings is better when those pairings have been learned via referent disambiguation (perhaps due to deeper processing) than through direct labeling, indicating that by 3 years old, referent disambiguation likely reflects deductive inference by the disjunctive syllogism (Zosh, Brinster, & Halberda, 2013).

A broader question is what 2.5-year-olds’ failure to apply the disjunctive syllogism to our tasks can tell us about their capacity for propositional thought in general. Based on their linguistic competence – for example, their use of abstract structure to interpret novel verbs (Fisher, 2002; Naigles, 1990), or even their ability to differentiate requests for “a truck or a ball” from “a truck and a ball” as described above – it seems unlikely that they lack the capacity for abstract, sententially-structured, propositional thought, since such thought appears to underlie their comprehension and production of language. Yet it has also been argued that 2-year-olds’ language production and comprehension largely reflects specific item-based knowledge, rather than abstract, rule-governed combination (e.g. Tomasello, 2000). This underscores the need to develop simpler tasks to better evaluate the logical inference-making abilities of very young children, and to integrate language measures in further studies of the development of logical reasoning.
References


use positive, but not negative, auditory cues to infer food location. *Animal Cognition, 15*(1), 45-55.


lion-tailed macaques (\textit{Macaca silenus}), a hamadryas baboon (\textit{Papio hamadryas}), capuchins (\textit{Sapajus apella}), and squirrel monkeys (\textit{Saimiri sciureus}). \textit{Journal of Comparative Psychology, 129}(3), 256–267.


Morris, B. J. (2008). Logically speaking: Evidence for item-based acquisition of the


