Cognition 154 (2016) 40–48

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

Original Articles

The emergence of reasoning by the disjunctive syllogism in early childhood

Shilpa Mody *, Susan Carey

Department of Psychology, Harvard University, Cambridge, United States

ARTICLE INFO

Article history:
Received 19 November 2014
Revised 11 May 2016
Accepted 16 May 2016
Available online 28 May 2016

Keywords:
Logic
Reasoning
Inference
Cognitive development
Disjunctive syllogism
Reasoning by exclusion

ABSTRACT

Logical inference is often seen as an exclusively human and language-dependent ability, but several nonhuman animal species search in a manner that is consistent with a deductive inference, the disjunctive syllogism: when a reward is hidden in one of two cups, and one cup is shown to be empty, they will search for the reward in the other cup. In Experiment 1, we extended these results to toddlers, finding that 23-month-olds consistently approached the non-empty location. However, these results could reflect non-deductive approaches of simply avoiding the empty location, or of searching in any location that might contain the reward, rather than reasoning through the disjunctive syllogism to infer that the other location must contain the reward. Experiment 2 addressed these alternatives, finding evidence that 3- to 5-year-olds used the disjunctive syllogism, while 2.5-year-olds did not. This suggests that younger children may not easily deploy this logical inference, and that a non-deductive approach may be behind the successful performance of nonhuman animals and human infants.

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.[1]

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

* Corresponding author at: 33 Kirkland St, Cambridge, MA 02138, United States.
E-mail address: shilpa@wjh.harvard.edu (S. Mody).

[1] 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 evaluated – this is sketched out in further footnotes.

http://dx.doi.org/10.1016/j.cognition.2016.05.012
0010-0277/

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 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, Koepke, Livingston, Girard, & Hartsfield, 2013; Schloegl, Schmidt, Boeckle, Weiß, & Kotrschal, 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 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".[2]

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 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.[3]

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.[4]

The cups task is ambiguous in regards to these three alternatives; using any 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 et al., 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

[2] 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.

[3] 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").

[4] 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.

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 and 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 2 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 s, the experimenter encouraged them to search for the ball until they approached one of the buckets or approximately 10 s elapsed. If they did not approach a bucket within 10 s, 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 ($t(23) = 6.362$, $p < 0.001$). 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) ($t(23) = 0.238$, $p = 0.812$). Furthermore, the toddlers were marginally successful on their first trial (sign test, 17/24 correct choices, $p = 0.064$), suggesting that they reason by exclusion spontaneously and without training.

These results establish that 23-month-olds, like older children, 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 (Fig. 1).

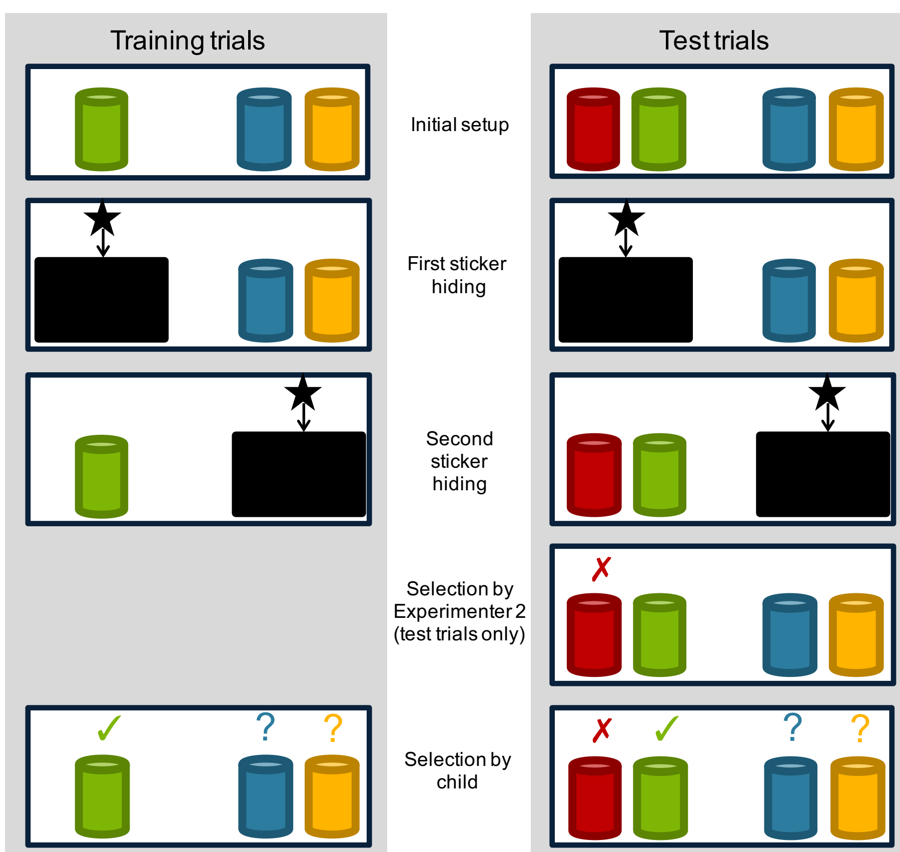

Fig. 1. Structure of training trials (left) and test trials (right) in Experiment 2. Symbols above the cups indicate the information that is available to the participant: the cup with the cross is empty, the cup with the checkmark is certain to contain a sticker, and the cups with question marks may or may not contain stickers.

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").

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 (Fig. 1) 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 (Fig. 1) 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 (Fig. 2).

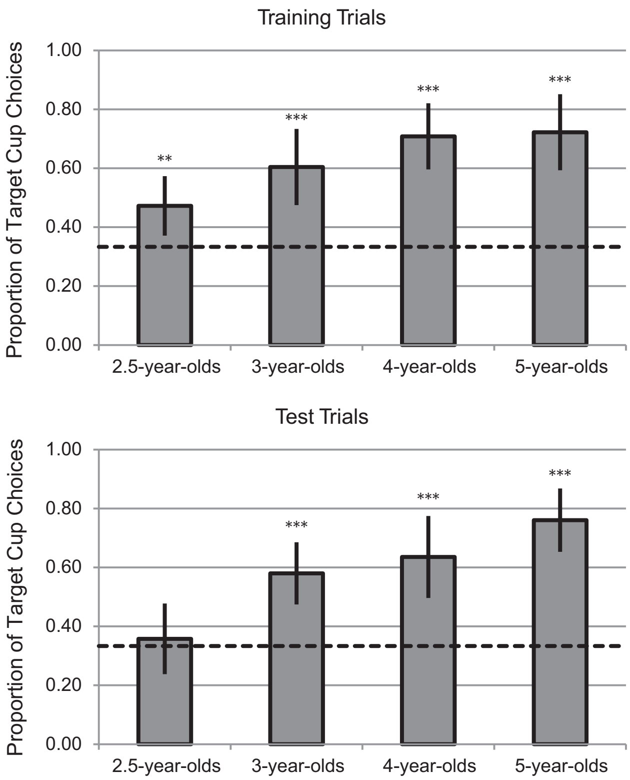

Fig. 2. Proportion of training trials (top) and test trials (bottom) in which children in each age group selected the target cup in Experiment 2. Error bars represent 95% confidence intervals, and the dotted line indicates chance (0.33).

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 0.0125. 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 ($t(23) = 2.853$, $p = 0.009$), 3-year-olds on 60% of trials ($t(23) = 4.339$, $p < 0.001$), 4-year-olds on 71% of trials ($t(23) = 6.918$, $p < 0.001$), and 5-year-olds on 72% of trials ($t(23) = 6.239$, $p < 0.001$). 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 et al. (2012), demonstrating that young 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%.[5] 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.[6]

As in the analysis of training trials, we used the Bonferroni correction for multiple comparisons, resulting in an adjusted alpha of 0.0125. 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 ($t(23) = 4.838$, $p < 0.001$), 4-year-olds on 64% of trials ($t(23) = 4.496$, $p < 0.001$), and 5-year-olds on 76% of trials ($t(23) = 8.225$, $p < 0.001$). 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 ($t(23) = 0.425$, $p = 0.675$). They behaved in the manner predicted by the non-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) ($t(95) = 0.818$, $p = 0.415$). 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 ($F(3,92) = 10.082$, $p < 0.001$), but no effect of trial type ($F(1,92) = 1.420$, $p = 0.236$), and no significant interaction ($F(3,92) = 0.814$, $p = 0.489$) 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 ($r(96) = 0.316$, $p = 0.002$) and test trials ($r(96) = 0.437$, $p < 0.001$). 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 ($r(94) = 0.374$, $p < 0.001$). 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

[5] 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.

[6] 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.

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.

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

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. Horobin & 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, 1985; 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.

Acknowledgements

We thank Jesse Snedeker, Roman Feiman, and Jean-Rémy Hochmann for invaluable comments and discussion, as well as Allison Hyland, Madison Singell, Evelyn Xing, Haven Jones, and Luke Anderson for help with data collection. This research was supported by NIH Grant #RO1-HD038338 to S.C.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.cognition.2016.05.012.

References

Andrews, G., & Halford, G. S. (2011). Recent advances in relational complexity theory and its application to cognitive development. In P. Barrouillet & V. Gaillard (Eds.), *Cognitive development and working memory: A dialogue between neo-piagetian and cognitive approaches* (pp. 47–68). Hove, East Sussex: Psychology Press.

Austin, K., Theakson, A., Lieven, E., & Tomasello, M. (2014). Young children's understanding of denial. *Developmental Psychology, 50*(8), 2061–2070.

Bloom, L., Lahey, M., Hood, L., Lifter, K., & Fiess, K. (1980). Complex sentences: Acquisition of syntactic connectives and the semantic relations they encode. *Journal of Child Language, 7*, 235–261.

Braine, M. D. S. (1978). On the relation between the natural logic of reasoning and standard logic. *Psychological Review, 85*(1), 1–21.

Braine, M. D. S., Reiser, B. J., & Rumain, B. (1984). Some empirical justification for a theory of natural propositional logic. *Psychology of Learning and Motivation, 18*, 313–371.

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

Call, J. (2006). Inferences by exclusion in the great apes: The effect of age and species. *Animal Cognition, 9*(4), 393–403.

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

Carruthers, P. (2002). The cognitive functions of language. *Behavioral and Brain Sciences, 25*, 657–674.

Cultice, J. C., Somerville, S. C., & Wellman, H. M. (1983). Preschoolers' memory monitoring: Feeling-of-knowing judgments. *Child Development, 54*(6), 1480–1486.

Erdohegyi, A., Topal, J., Viranyi, Z., & Miklosi, A. (2007). Inferential reasoning in a two-way choice task and its restricted use. *Animal Behavior, 74*, 725–737.

French, L. A., & Nelson, K. (1985). *Young children's knowledge of relational terms: Some ifs, ors, and buts*.New York: Springer-Verlag.

Heimbauer, L. A., Antworth, R. L., & Owren, M. J. (2012). Capuchin monkeys (*Cebus apella*) use positive, but not negative, auditory cues to infer food location. *Animal Cognition, 15*(1), 45–55.

Hill, A., Collier-Baker, E., & Suddendorf, T. (2011). Inferential reasoning by exclusion in great apes, lesser apes, and spider monkeys. *Journal of Comparative Psychology, 125*, 91–103.

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

Horobin, K., & Acredolo, C. (1989). The impact of probability judgments on reasoning about multiple possibilities. *Child Development, 60*(1), 183–200.

Johnson-Laird, P. N. (2010). Mental models and human reasoning. *Proceedings of the National Academy of Sciences, 107*(43), 18243–18250.

Johnson-Laird, P. N., Byrne, R. M., & Schaeken, W. (1992). Propositional reasoning by model. *Psychological Review, 99*(3), 418–439.

Lea, R. B. (1995). On-line evidence for elaborative logical inferences in text. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(6), 1469–1482.

Lust, B., & Mervis, C. A. (1980). Development of coordination in the natural speech of young children. *Journal of Child Language, 7*, 279–304.

Lyons, K. E., & Ghetti, S. (2011). The development of uncertainty monitoring in early childhood. *Child Development, 82*(6), 1778–1787.

Maille, A., & Roeder, J. J. (2012). Inferences about the location of food in lemurs (*Eulemur macaco* and *Eulemur fulvus*): A comparison with apes and monkeys. *Animal Cognition, 15*(1), 1075–1083.

Mikolasch, S., Kotrschal, K., & Schloegl, C. (2012). Is caching the key to exclusion in corvids? The case of carrion crows (*Corvus corone corone*). *Animal Cognition, 15*(1), 73–82.

Mikolasch, S., Kotrschal, K., & Schloegl, C. (2011). African grey parrots (*Psittacus erithacus*) use inferences by exclusion to find hidden food. *Biology Letters, 7*(6), 875–877.

Miscione, J. L., Marvin, R. S., O'Brien, R. G., & Greenberg, M. T. (1978). A developmental study of preschool children's understanding of the words "know" and "guess". *Child Development, 49*(4), 1107–1113.

Moore, C., Bryant, D., & Furrow, D. (1989). Mental terms and the development of certainty. *Child Development, 60*(1), 167–171.

Morris, B. J. (2008). Logically speaking: Evidence for item-based acquisition of the connectives AND & OR. *Journal of Cognition and Development, 9*(1), 67–88.

Morris, B. J., & Sloutsky, V. (2002). Children's solutions of logical versus empirical problems: What's missing and what develops? *Cognitive Development, 16*, 907–928.

O'Brien, D. P., Braine, M. D., & Yang, Y. (1994). Propositional reasoning by mental models? Simple to refute in principle and in practice. *Psychological Review, 101*(4), 711–724.

Osherson, D. N., & Markman, E. (1975). Language and the ability to evaluate contradictions and tautologies. *Cognition, 3*(3), 213–226.

Paukner, A., Huntsberry, M. E., & Suomi, S. J. (2009). Tufted capuchin monkeys (*Cebus apella*) spontaneously use visual but not acoustic information to find hidden food items. *Journal of Comparative Psychology, 123*, 26–33.

Penn, D. C., Holyoke, K. J., & Povinelli, D. J. (2008). Darwin's mistake: Explaining the discontinuity between human and nonhuman minds. *Behavioral and Brain Sciences, 31*, 109–178.

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

Pillow, B. H., & Anderson, K. L. (2006). Children's awareness of their own certainty and understanding of deduction and guessing. *British Journal of Developmental Psychology, 24*(4), 823–849.
Premack, D. (2007). Human and animal cognition: Continuity and discontinuity. *Proceedings of the National Academy of Sciences, 104*(35), 13861–13867.
Premack, D., & Premack, A. J. (1994). Levels of causal understanding in chimpanzees and children. *Cognition, 50*(1), 347–362.
Rescorla, M. (2009). Chrysippus' dog as a case study in non-linguistic cognition. In R. W. Lurz (Ed.), *The philosophy of animal minds* (pp. 52–71). Cambridge, UK: Cambridge University Press.
Rips, L. J. (1994). *The psychology of proof: Deduction in human thinking*.Cambridge, MA: MIT Press.
Russell, J., & Haworth, H. M. (1987). Perceiving the logical status of sentences. *Cognition, 27*, 73–96.
Sabbatini, G., & Visalberghi, E. (2008). Inferences about the location of food in capuchin monkeys (*Cebus apella*) in two sensory modalities. *Journal of Comparative Psychology, 122*, 156–166.
Schloegl, C., Dierks, A., Gajdon, G. K., Huber, L., Kotrschal, K., & Bugnyar, T. (2009). What you see is what you get? Exclusion performances in ravens and keas. *Plos One, 4*, 1–12.
Schloegl, C., Schmidt, J., Boeckle, M., Weiß, B. M., & Kotrschal, K. (2012). Grey parrots use inferential reasoning based on acoustic cues alone. *Proceedings of the Royal Society B, 279*(1745), 4135–4142.
Schmitt, V., & Fischer, J. (2009). Inferential reasoning and modality dependent discrimination learning in olive baboons (*Papio hamadryas anubis*). *Journal of Comparative Psychology, 123*, 316–325.
Spelke, E. (2002). Developing knowledge of space: Core systems and new combinations. In S. Kosslyn & A. Galaburda (Eds.), *Languages of the brain* (pp. 239–258). Cambridge, MA: Harvard University Press.
Watson, J. S., Gergely, G., Csanyi, V., Topal, J., Gasci, M., & Sarkozi, Z. (2001). Distinguishing logic from association in the solution of an invisible displacement task by children (*Homo sapiens*) and dogs (*Canis familiaris*): Using negation of disjunction. *Journal of Comparative Psychology, 115*, 219–226.
Zelazo, P. D., & Frye, D. (1998). Cognitive complexity and control II. The development of executive function in childhood. *Current Directions in Psychological Science, 7*(4), 121–126.