

Young children's flexible use of semantic cues to word meanings: converging evidence of individual and age differences*

GEDEON O. DEÁK

University of California, San Diego

AND

GAYATHRI NARASIMHAM

Vanderbilt University

(Received 14 June 2010 – Revised 13 August 2012 – Accepted 13 December 2012 –
First published online 7 May 2013)

ABSTRACT

A new test of children's flexible use of semantic cues for word learning extended previous results. In Experiment 1, three- to five-year-olds ($N = 51$) completed two tests of interpreting several novel words for the same stimulus arrays. Within-sentence phrasal cues implied different stimulus referent properties. Children's cue-using flexibility in the new *Flexible Induction of Meanings [Words for Animates]* test (FIM-An) was strongly correlated with an established test (*Flexible Induction of Meanings [Words for Objects]*; Deák, 2000). Individual children showed between-test consistency in using cues to flexibly assign words to different referent properties. There were large individual differences, as well as limited age differences, in the distribution of flexible and inflexible response patterns. The comprehensibility of specific cues, and perceptual salience of specific properties, explained much of the variance. Proportions of flexible and inflexible patterns shifted with age. Experiment 2 replicated these results in $N = 36$ three- and four-year-olds, using a modified FIM-An with more distinctive cues.

[*] This research was supported by a Hellman fellowship and by National Science Foundation award BCS-0092027 to the first author. Thanks to the children who participated, and to Sarah Creel, Anna Holt, and Jenny Nowinski for comments on an earlier draft. Address for correspondence: Gedeon O. Deák, Dept. of Cognitive Science, 9500 Gilman Dr., University of California, San Diego, La Jolla CA 92093-0515, USA.

INTRODUCTION

Children, as well as adults, use linguistic context to interpret unfamiliar words or phrases that they hear in conversation or read in text (Beck, McKeown & McCaslin, 1983; Nagy, Anderson & Herman, 1987; Thorndyke, 1976). In school-aged children, this ability strongly predicts reading skill (Cain & Oakhill, 1999; Cain, Oakhill & Lemmon, 2004; Levy, Abello & Lysynchuk, 1997). Children's ability to utilize contextual cues might require COGNITIVE FLEXIBILITY, or the capacity to update attention, representation, and inference, in response to changing task-relevant information. Cognitive flexibility might play a critical role in word learning and language comprehension. For example, in order to understand ongoing conversation or narratives, listeners must update their representations of a speaker's meanings 'on the fly' (Cain *et al.*, 2004; Deák, 2003).

A standing question is how children learn to use context for comprehension, even before they start school. Do preschool-aged children show consistent (i.e., between-task) ability to utilize contextual cues to word meanings? It has been shown repeatedly that preschool-aged children gradually learn to use a wide range of contextual cues to infer word meaning (e.g., Landau, Smith & Jones, 1998; Smith & Yu, 2008). Moreover, there is some evidence of cross-situational consistency in individual school-aged children's tendency to use contextual cues (e.g., Gray, 2004). If preschoolers also show consistent individual differences, these differences might help us predict which children are likely to have similar difficulties in reading comprehension. However, there are almost no data on individual preschool children's tendencies to use different cues to infer different word meanings. Also, there are almost no data on individual differences relative to age differences in preschool-aged children.

To address this, the current study focused on young children's use of contextual cues to flexibly infer the meanings of several words. Because cue use differs across problems, we consider linguistic and non-linguistic content that might contribute to differences. Because individual children might show similar or different patterns of cue use across problems or tests, we focused on two possible contributing factors: children's comprehension of verbal cues, and children's ability to flexibly shift inferences about meaning.

Children's use of sentential cues to word meaning

Preschool children, like adults, can interpret novel words in light of contextual cues of many types, from morphological to paralinguistic (e.g., Akhtar, Carpenter & Tomasello, 1996; Saylor, Sabbagh & Baldwin, 2002; Storkel, 2001; Ellis Weismer & Hesketh, 1993; Yu & Ballard, 2007). This ability improves with age, as children learn more diverse and subtle

meanings and constructions (e.g., Arnold, Brown-Schmidt & Trueswell, 2007). However, most previous studies have examined one or two cues at a time, within a single task context. Thus, as much as we have learned from these studies, questions remain about how children negotiate the complex, variable, and changing contexts of the many novel words they encounter.

Among the important cues to a novel word's meaning are the words and phrases surrounding it. Even two-year-olds can, in optimal tasks, use meaningful within-sentence content to interpret novel words (e.g., Goodman, McDonough & Brown, 1998; Samuelson & Smith, 1999). However, meaningful cues are not only *INTRA*-sentential (e.g., the verb of a novel noun) but also *INTER*-sentential (e.g., the topic of conversation; Umstead & Leonard, 1983). Inter- and intra-sentential cues differ in *SCOPE*, and the integration of cues over different scopes can present a challenge. Scope refers to the range of influence of a cue upon another element. We refer to cues like a verb's inflection or the article of a noun as 'local' scope cues. More extended information, such as the prevailing topic of conversation, are 'distal' scope cues. Children aged three to five years can sometimes synthesize cues across clauses and sentences (Arnold *et al.*, 2007), but they also make 'scope errors'. For example, they may ignore thematic information from recent sentences, when inferring the intended meaning of a homophone (Campbell & Bowe, 1983). Children also make the opposite error: they may allow 'weak' cues from previous sentences to inappropriately govern their inferences about a word's meaning, even if within-sentence cues provide contradictory information (Campbell & Bowe, 1983; Deák, 2000). Such errors suggest a long process of learning what cues, distant and local, determine the meaning of various kinds of novel words.

As noted, a novel word's meaning is crucially constrained by its predicate or other semantically linked intra-sentential phrases. For example, compare: 'The [WORD₁] in the tree held a branch', 'The monkey in the [WORD₂] held a branch', and 'The monkey in the tree held a [WORD₃]'. The meaningful phrases that are present in each version, and their structural relation to the novel word, strongly imply different meanings. Adults can keep the different implications and contexts straight, and distinguish between the likely meanings of the words, even if the sentences occur in succession. Yet when young children hear several novel words in succession, each with different intra-sentential cues, they sometimes make scope errors. That is, they inappropriately allow previous (inter-sentential) cues to trump the intra-sentential cues. For example, a previous phrase might dictate a child's inference about the current word.

The cause of these errors remains unclear. They might reflect children's slow accrual of semantic and conceptual knowledge. When children only weakly understand the current intra-sentential cue, they might refer back to previous cues that they understood more clearly. It is known that children's

comprehension of a cue context affects their inferences about unfamiliar words or constructions. For example, children who are experts in a given domain tend to make relatively sophisticated inferences about novel words in that domain (Johnson & Mervis, 1994). However, we know little about how children more generally use ‘local’ cues to interpret novel words from a variety of domains.

Children’s flexible use of changing cues

Another possible factor in children’s use of phrase cues is cognitive control. In dynamic discourse contexts (e.g., changing cues, structures, topics, etc.) children must continually select some cues while ignoring others, to update their understanding of a speaker’s successive utterances. This adaptive selection and interpretation of changing cues entails COGNITIVE FLEXIBILITY (Deák, 2003). Cognitive flexibility¹ might affect children’s use of intra-sentential phrase cues, over and above their degree of comprehension of the cues. Flexibility might especially matter when children hear several new words that are associated with an ongoing or recurrent topic. These situations require children to update their selection and interpretation of verbal cues, sometimes when there is conflict from earlier cues or inferences. Flexibility and related cognitive control skills develop markedly during early childhood (Davidson, Amso, Anderson & Diamond, 2006). Children’s scope errors might partly stem from their limited flexibility under circumstances of conflict, for example, when cues and focus words are changing over a series of comments about a topic.

To investigate how flexibility relates to word learning, we designed a new test in which children must use changing phrase cues to interpret several words for the same stimulus array (or ‘topic’). We predicted age differences as well as individual differences in flexibility. We also investigated whether phrase cues could override children’s prior bias towards certain interpretations – that is, their tendency to favor certain stimulus properties or meanings above others. If some meanings are more compelling than others, this might pose a challenge to flexibility (Ellefsen, Shapiro & Chater, 2006) over and above the challenge imposed by changing and conflicting cues. Furthermore, when earlier cues favor a privileged meaning, children might especially tend to persist in assigning that meaning to later words (i.e., scope errors).

There is ample evidence that young children rapidly develop flexibility in responding to verbal cues. Most of the evidence is from rule-switching tasks. In these tasks, children sort two cards (e.g., orange rabbit; blue boat)

[1] *Flexibility* means selectively encoding, interpreting, synthesizing, and de-selecting cues in changing tasks.

into two boxes (Dimensional Change Card Sorting test; Zelazo, 2006). They first follow one dimension-based rule (e.g., color) and then are re-instructed to sort the cards using the other feature rule (i.e., shape). This requires a reversal of the sorting 'policy'. These tasks show robust age-related changes (Cepeda, Kramer & De Sather, 2001; Zelazo, Reznick & Piñon, 1995). Many three-year-olds persevere by continuing to use the first rule after being told to switch rules. Fewer four-year-olds make these perseverative scope errors. This is evidence for developmental changes in flexibility, but it is not clear that these changes extend to word-learning processes. There are fundamental differences between following rules and inferring word meanings, related to demands for flexibility. For example, rule-instructions are definite, repetitive, and imperative, whereas word-meaning cues are diverse, implicit, and probabilistic. In card-sorting tests, children repeatedly choose between two simple responses to the same two rules for the same two stimuli. By contrast, word meanings are indefinitely broad, with a wide range of possible cues. Finally, the rule-switch is arbitrary and unmotivated, whereas word meanings do not arbitrarily switch or rearrange. Given these differences, it is not clear that age differences in rule-switching flexibility will generalize to word-learning flexibility.

Nonetheless, there is evidence that flexible word learning develops from three to four years, the same period in which rule switching develops. During this time children improve at using phrases to select relevant properties of complex stimuli (Kalish & Gelman, 1992). For example, in studies using the *Flexible Induction of Meanings for Object-Words* test (FIM-Ob; Deák, 2003; Deák & Narasimham, 2003), preschool children see sets of objects with various shapes, materials, and parts. Children hear three novel words for each set, and each word follows a different predicate phrase: 'is a ...', 'is made of ...', and 'has a ...'. The 'is a' phrase implies a shape-based category (Baldwin, 1989; Landau *et al.*, 1998); 'made of' implies a material-kind category (Dickinson, 1988), and 'has a' implies a salient part (Saylor *et al.*, 2002). To be flexible, children must adapt their interpretations of each successive word to the current phrase cues, while ignoring previous cues and previous responses.

In the FIM-Ob test, accuracy in earlier AND later blocks increases with age: almost all five- and six-year-olds flexibly adapt to all three phrase cues, whereas many three-year-olds use the first cue to interpret the first word, but then persevere, assigning the same meaning to the later words.² In addition, other adaptive and maladaptive patterns of inferences across

[2] This is a peculiar error: imagine a mechanic explaining that your car's problem is that 'the rotor has a bad caliper,' and the new one should be 'made of chrome-plated steel.' It would be odd to interpret all three words as labels for the same object. Yet this is analogous to some children's errors in the FIM-Ob.

trials become apparent (because the FIM is more complex and therefore more sensitive than traditional, binary card-sorting tests): some children are indeed consistently flexible; others, however, are only partly flexible (e.g., they utilize only two of the three phrase cues). Some children perseverate on a single dimension (e.g., material); others perseverate on a favored object within each set, but not based on any common feature. Other children are 'indiscriminate,' switching inferences (i.e., not perseverating), but not based on any clear cue or principle. The incidence of these patterns changes with age: most three-year-olds perseverate or are indiscriminate, many four-year-olds are partly flexible or perseverative, and most five-year-olds use all of the cues flexibly. One interpretation is that perseverative errors represent 'classic' inflexibility, wherein prior biases or information are not ignored, whereas indiscriminate errors stem from limited comprehension, because children seem not to derive the implications of the phrase cues. If these interpretations are valid, then the shifting proportions of these response-patterns from three to five years would suggest that children's adaptive use of cues to infer word meanings develops as a result of both increasing cue comprehension and increasing cognitive flexibility.

This interpretation is, however, limited by the fact that the FIM-Ob test examined only three cues ('is a', 'is made of', 'has a') for three properties (shape, material, and part) of one stimulus kind (artifact-objects). We do not know whether there are similar age and individual differences with other cues, properties, or stimulus kinds. This leaves open a question of convergent validity (Campbell & Fiske, 1959). For example, we do not know whether age and individual differences in the FIM-Ob test generalize to words for properties of biological kinds. By developing multiple tests of flexibility, with different cues, properties, and kinds, we can better understand how children's growing semantic knowledge (i.e., sentence comprehension) and cognitive flexibility contribute to their ability to use changing cues to infer the meanings of several words for a topic.

We designed a new assessment, the *Flexible Induction of Meanings for Words-Animates* test, or FIM-An. Children see sets of colorful drawings of novel creatures or aliens, holding unfamiliar possessions, situated in strange environments. Within each set there is a STANDARD item and four comparison items. Three of these comparison items each share a single property with the standard: one shows the same alien 'species', one shows the same habitat, and one shows a creature with the same possessed, or held, object. A word presented in reference to the standard could refer to one of these properties, and therefore generalize to one of the comparison items. A fourth comparison item differs in all three properties, as a means to check for guessing or inattention. Figure 1 shows two of the four stimulus sets.

There were several reasons for choosing animals as a domain to investigate children's novel word inferences. Animals are complex and

WORD-LEARNING FLEXIBILITY IN CHILDREN

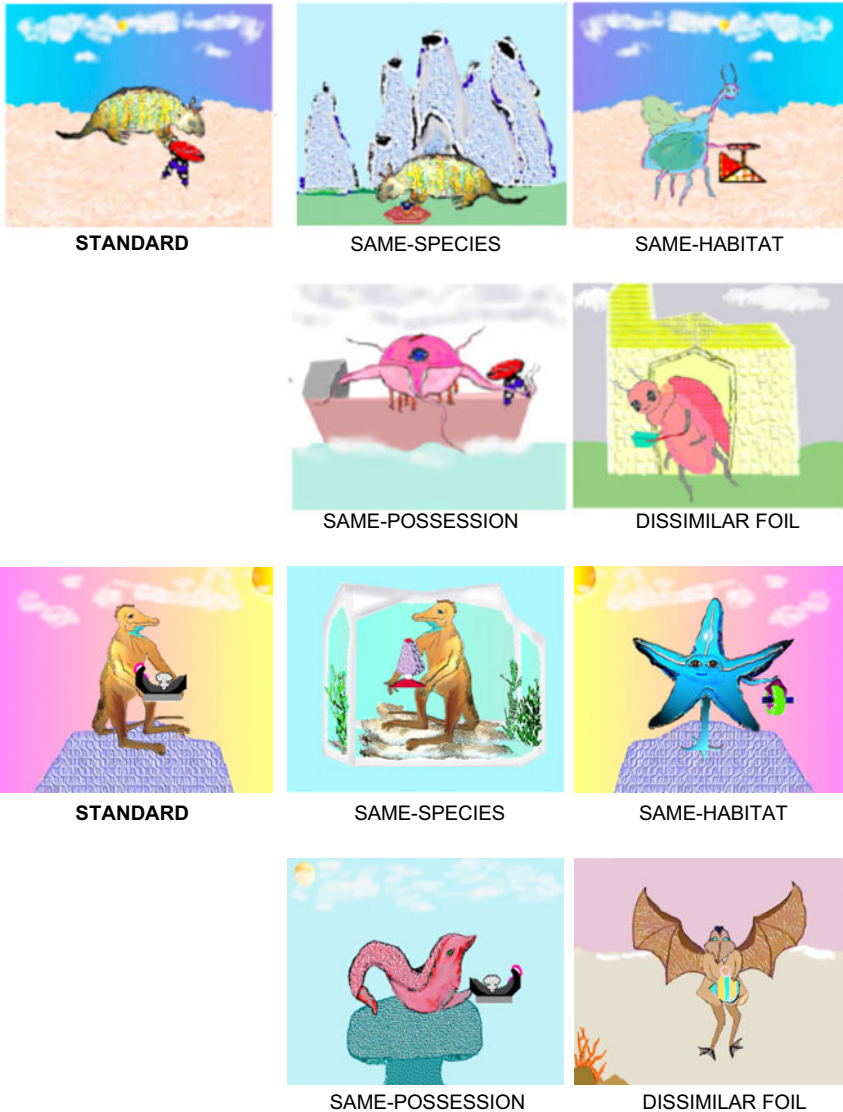


Fig. 1. (Colour online) Examples of two stimulus set from the FIM-An tests. Pictures in each set are, clockwise from upper left: standard, same-species, same-habitat, dissimilar foil, and same-possessed-object items.

typically have a wide range of labeled properties. Children hear many words for properties of animals while reading books, visiting zoos, caring for pets, or watching nature shows. Also, children are generally interested in

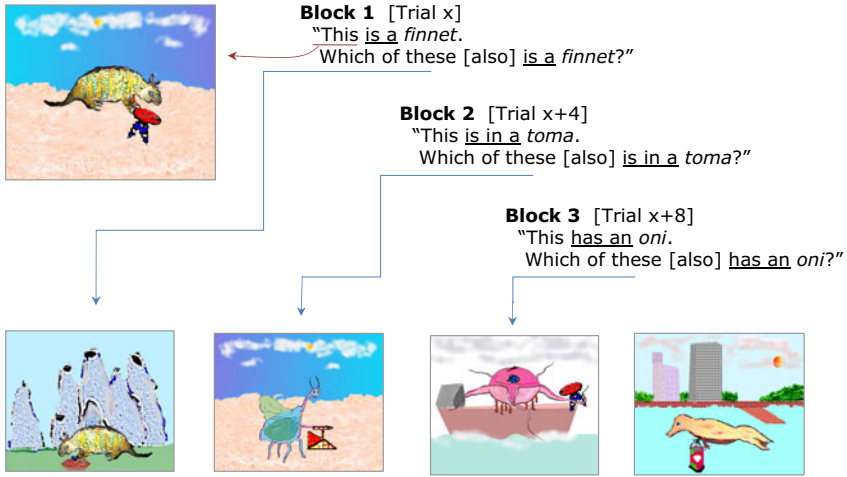


Fig. 2. (Colour online) Example of relations of phrase cues, novel words, and possible named stimulus properties. In this example, children hear different words following the phrase cues (in succession) 'is a ...', 'is in a ...', and 'has a ...'. Arrows indicate cue-appropriate inductive choice.

animals, their physical traits, behaviors, environments, etc. (DeMarie, 2001; Tomkins & Tunnicliffe, 2007). Finally, animals have different properties, property labels, and predicates than objects. Thus, by comparing the new test to the FIM-Ob test, we can test the generality of children's word-learning flexibility across conceptual domains.

In our first exploratory experiment, we tested children's flexible adaptation to phrase cues that imply different properties. The predicate phrase 'is a ...' was expected to imply a species relation; 'is in [on] a ...' was expected to imply a habitat; and 'has a ...' was expected to imply a possessed object.

In the FIM-An test, children saw each of four sets of pictures three times. Each time they heard a different novel word following a different phrase cue. Children could then generalize each word to a same-species, same-habitat, or same-possession item. The design of the test is shown in Figure 2.

The FIM-An test is comparable to the FIM-Ob test (Deák, 2000). Children's responses can be evaluated with respect to specific cues (e.g., same-habitat interpretations of words following 'is [in/on] a ...'), or in terms of flexibility. Flexibility was estimated from children's ratios of different inferences about different words for the same stimulus sets, based on the changing phrase cues. In addition, in order to estimate how specific cues facilitated flexibility, we used children's responsiveness to a given cue when

it was the first cue for a stimulus set (rather than the second or third). That is, first-block accuracy was an estimate of children's baseline ability to select, understand, and utilize each cue. This baseline can be compared to later trials, when there are added demands for flexibility. This can reveal cue-specific limitations on cue use: for example, if children are unclear about the meaning of a given cue, they might perseverate on previous inferences or respond haphazardly when the cue follows other inferences. This is important because not every predicate cue has equal strength (i.e., tendency to imply, for fluent language-users, a specific, transparent referent-property). It is unknown how changes in cue strength from earlier to later cues might impact flexibility. By assessing the baseline strength of every cue in every test, we can estimate this effect. Also, by adjusting individual children's flexibility scores based on the estimated strength of earlier and later cues (at a group level), flexibility can be compared across tests and children. This provides improved estimates of individual differences in word-learning flexibility.

Experiment 1 tested children's use of INITIAL phrase cues and LATER phrase cues to infer successive word meanings in the FIM-An and FIM-Ob tests. Experiment 2 explored a modified FIM-An test, with more stimulus sets (for greater reliability), and more semantically explicit phrase cues.

EXPERIMENT 1

The main goal of Experiment 1 was to examine correspondences between the FIM-An and FIM-Ob tests. We predicted age and individual differences in semantically appropriate cue use, both when that cue was first for a stimulus set, and when it followed other cues and inferences.

In addition to these differences, the cues in both tests probably vary in strength. Deák (2000) found that in the FIM-Ob test, 'is made of ...' implies a material category more strongly than 'has a ...' implies a part, and 'is a ...' (or 'looks like a ...') implies an object shape rather weakly. Still, most four-year-olds adapt to all three cues to some degree, so the cues are age-appropriate. Similarly, we expected the FIM-An test cues to vary in strength, although most children (at least four-year-olds) should understand the cues to some degree.

The FIM-An test examined children's use of three different phrase cues to assign words to different properties of animate entities. First, the phrase 'is a' followed by a bare count noun implies a generic category label. This implication is strong for natural kinds (Gelman, 2003). For example, parents often use the 'is a' copula to label natural kinds for children (Gelman, Coley, Rosengren, Hartman, Pappas & Keil, 1998). Conversely, children infer 'deep' properties for creatures labeled in this way (Gelman & Heyman, 1999; Gelman *et al.*, 1998). Also, the FIM-An test same-species

stimuli displayed physical properties that preschoolers use to categorize animal kinds (see, e.g., Jones, Smith & Landau, 1991). Thus, children should generalize a noun following ‘is a ...’ to the same-species stimulus.

Second, with respect to ‘is [in/on] a ...’, two-year-olds interpret ‘in’ and ‘on’ as denoting canonical spatial contexts (Corrigan, Halpern, Aviezer & Goldblatt, 1981; Grieve, Hoogenraad & Murray, 1977). Also, by four years, children talk about animals’ habitats in response to the probe, ‘What lives in a forest?’ (Huxham, Welsh, Berry & Templeton, 2006; Strommen, 1995). Finally, same-habitat items had redundant features to support a spatial-context match; and the prepositions ‘on’ or ‘in’ were chosen to fit each set’s details. Thus, young children relate ‘in’ or ‘on’ to habitats, and know that habitats have labels.

Third, children know that ‘have’ or ‘got’ imply possession: about half of children aged 2;6 to 3;0 spontaneously use those verbs to describe possession relations (Friedman & Neary, 2008; Hay, 2006). Deák, Ray and Brenneman (2003) found that four-year-olds tended to answer the questions ‘What does it have?’ by naming something an animal is holding rather than the animal itself. Thus, most young children are aware that ‘have’/‘got’ can imply possessions, which can be shown, canonically, as objects that are held or worn by an individual.

Despite this evidence that children should understand all three phrases, the phrases might differ in strength. For example, children might readily generalize putative nouns to species, because they expect nouns for animals to refer to categories roughly like species (Gelman, 2003), and because the creatures are relatively salient in the stimuli. If this (or any) cue is stronger than the others, it might cause cue-order effects. That is, children might have more trouble shifting from species-based to habitat- or possession-based interpretations of later words, than the reverse. Although it remains unclear whether it is generally harder to switch to stronger tasks or to weaker tasks (Deák, Ray & Pick, 2004; Ellefson *et al.*, 2006), the question clearly pertains to language processing, where the strength of successive cues will rise and fall unpredictably. In order to explore how these perturbations affect children’s comprehension, cue order in this investigation was counterbalanced. This allowed us to test whether cue order – especially the first cue – affected the incidence of later flexible or perseverative patterns.

One consequence of this design is that different children had different cue orders, making it more difficult to assess individual differences. Although there is no perfect way to measure both cue-order effects and individual differences within the same design, we describe below a general strategy for comparing individual children’s flexibility across different tasks or task orders. Briefly, we assign ‘credit’ for each cue-appropriate switch in later

trials, as a proportion of all possible switching opportunities. Credit for switches can be weighted according to the relative baseline strength of each successive cue, in order to control for cue order.

To assess receptive language, children completed the Peabody Picture Vocabulary Test (Dunn & Dunn, 1981). As a measure of verbal inhibition, or ignoring known word–referent associations, children completed the Stroop Day–Night Test (Gerstadt, Hong & Diamond, 1994).

METHOD

Participants

Fifty-three three- and four-year-old children were recruited and tested in preschools in Nashville, TN.³ All children were fluent in English, and most were Caucasian and middle class. Two children were excluded because their age-standardized PPVT-R scores were two *SDs* below average. The remaining fifty-one children averaged 48.3 months of age (range = 37 to 59), and included twenty-six girls (mean age = 47 months) and twenty-five boys (mean = 49 months). PPVT-R scores showed age-typical receptive language skills among girls (mean = 105.0, *SD* = 11.8) and boys (mean = 101.6, *SD* = 10.9).

Materials

The FIM-An test used four sets of five computer-drawn color pictures (12.5 cm × 15 cm), each showing a novel creature (inspired by Barlow & Summer, 1979), in an unfamiliar environment, holding a novel object. Each set included a standard picture and four comparison pictures. Thus, sixteen novel species, sixteen habitats, and sixteen held objects were presented in twelve trials (three trials per set). The standards were labeled with twelve distinct bisyllabic neologisms or rare words (e.g., *eland*, *finell*, *indri*, *minar*, etc.).

The FIM-Ob test used six sets of five physical objects, also with a standard and four comparison objects per set (Deák & Narasimham, 2003). Each standard had a novel shape, material, and part. Comparison objects included one same-shape object, one same-material object, one same-part object, and a distracter (see Figure 1). Features were named with sixteen different neologisms.

The PPVT-R used standardized picture plates (Dunn & Dunn, 1981). The Stroop Day–Night task used cards (13.5 × 10 cm) showing a sun in a blue field, or a moon and stars in a black field.

[3] Some results from the FIM-Ob in the full sample (see below) are described in Deák and Narasimham (2003).

Procedure

Children were tested individually over two sessions in a quiet room in their preschool. In the first session children completed the FIM-Ob test; in the second session they completed the FIM-An test. After the flexibility test in each session, children completed either the PPVT or Stroop Day–Night Test, counterbalanced between sessions. (There were no significant order effects.)

In the FIM-An test, each child saw four sets of novel pictures, once per block, in each of three blocks. On each trial the child was told to examine every picture. The experimenter then pointed to the standard, and stated and repeated a fact about it. Each fact had a novel word following one of three cues, ‘is a ...’, ‘is in [on] a ...’, or ‘has a ...’. The child was then asked to generalize the word to another picture (e.g., ‘This is a *finnel*. Which of these others is also a *finnel*?’). After the first block of trials, the adult presented the sets again in the same order, with different words and a new cue. The design is shown in Figure 3. Children were prompted after 8 seconds if they did not respond. They received non-specific verbal feedback (i.e., ‘Thank you’), delivered with consistent prosody, after every response.

The FIM-Ob test was administered as described by Deák and Narasimham (2003). The design was analogous to the FIM-An test sequence, except there were six sets and therefore six trials per block (total = 18 trials). On each trial, children examined the objects and then were told a fact about the standard, with a unique novel word and one of the phrase cues: ‘looks like a ...’, ‘is made of ...’, or ‘has a(n) ...’. (For example: ‘This one is made of *inrom*. Which of these also is made of *inrom*?’) In both flexibility tests, order of sets within blocks was randomized, but the order was repeated in every block so that the interval between presentations of a given set was constant. Words were randomized over trials for every child. Cue order was counterbalanced across blocks.

The Peabody Picture Vocabulary Test-Revised was administered using the standard procedure (Dunn & Dunn, 1981). On each trial the experimenter said a word, and the child pointed to the referent, shown as one of four pictures on a plate. Words gradually increased in difficulty. The Stroop Day–Night Test was administered as in Gerstadt *et al.* (1994): children were told to say ‘day’ and ‘night’, respectively, to simple pictures of the moon and the sun. After six practice trials with feedback, children completed sixteen test trials (eight per card) in quasi-random order, without feedback.

Responses in all tests were coded on-line. Videos were later coded off-line to check accuracy. Stroop response time (RT) was assessed through frame-by-frame coding of videos of each trial.

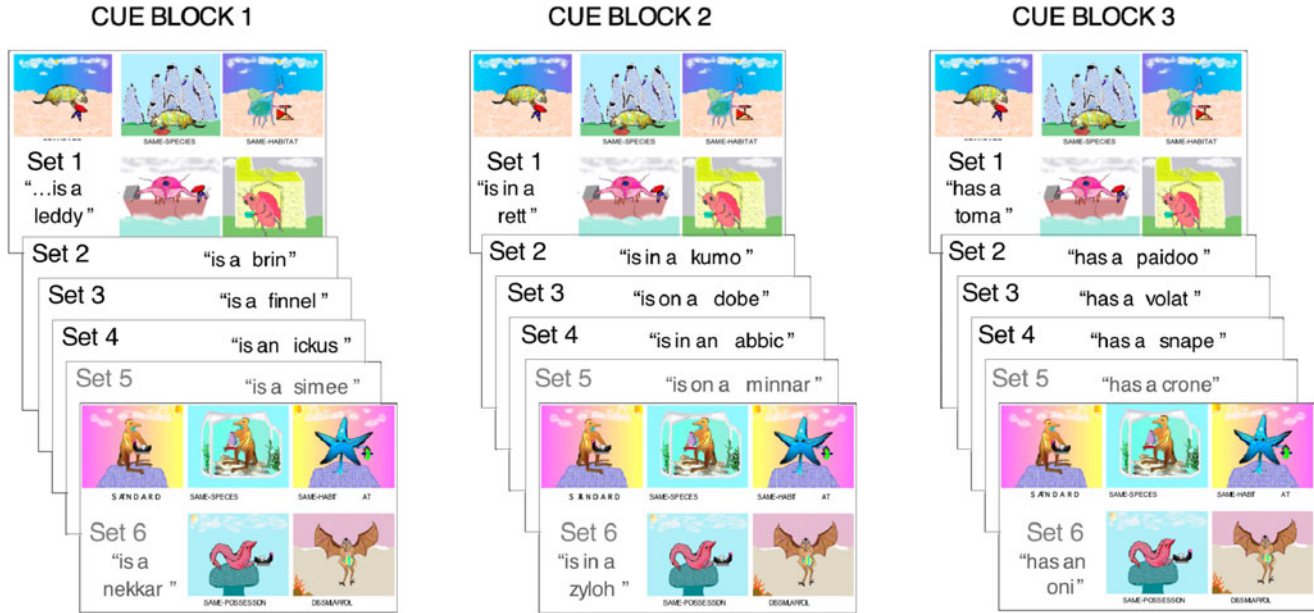


Fig. 3. (Colour online) Test sequence in FIM-An. In this example, children hear phrase cues in the order ‘is a ...’ in block 1, ‘lives in/on a ...’ in block 2, and ‘has a ...’ in block 3. The FIM-Ob uses the same scheme. See text.

FIM analyses

Proportion of appropriate responses for each cue can provide an estimate of children's comprehension of that cue. Flexibility can be estimated from the proportion of cue-appropriate responses in the second and third trial blocks. Specifically, we calculated the ratio of appropriate switches to the number of opportunities for an appropriate switch (CORSWOPS). CORSWOPS is conservative: children get 'credit' only for cue-appropriate inferences that differ from the previous inference. That is, if a child repeats an earlier choice that was not previously appropriate, but becomes appropriate after the cue changes, this is not counted as a flexible switch. Because it is a ratio, CORSWOPS is comparable across a wide variety of tests. However, it must meet the assumption that subjects make enough early cue-appropriate responses to provide numerous opportunities for flexible switches. Thus, for a more valid and conservative assessment of flexibility, we focused our analyses on children who responded correctly to many first-block cues. Although this makes our tests of flexibility more interpretable, it also means that our results cannot be considered age-typical, because some children were not included in the analyses.

In order to further control for baseline accuracy differences, in tests of flexibility, we used first-block accuracy as a covariate. We also re-ran all analyses using adjusted CORSWOPS values, wherein correct switches were weighted according to baseline cue strength. That is, cue-appropriate switches received more 'credit' if the previous cue was relatively strong and therefore more likely to generate subsequent perseverative responses; conversely, switches received less credit if the previous cue was weak, and thus unlikely to compel subsequent perseveration. Weightings were determined by group-level means for each cue. However, the results from these weighted scores did not differ from those described below, so we do not report them here. A full account is available by request from the first author.

In sum, to more accurately estimate the effects of cue comprehension and the effects of cognitive flexibility, our strategies (actually general strategies that can apply to any study of flexibility) were to: (1) assess baseline difference in cue use, and compare to post-switch accuracy; (2) use a normalized measure of flexibility, CORSWOPS, to adjust for each child's prior cue-appropriate responses, optionally weighting each appropriate switch according to the strength of the prior cue; and (3) use children's pre-switch accuracy as a covariate. These strategies were used in both experiments.

Response pattern definitions

Children produce one of four response patterns in the FIM-Ob test (Deák, 2000; Deák & Narasimham, 2003): flexible ($\geq 78\%$ appropriate choices; $\geq 67\%$ correct switches); partly flexible ($\geq 56\%$ appropriate; $\geq 33\%$ correct

switches); perseverative ($\leq 50\%$ appropriate; $\leq 25\%$ correct switches; $\leq 33\%$ total switches); or indiscriminate ($\leq 56\%$ appropriate; $\leq 25\%$ correct switches; $\geq 42\%$ total switches). These categories neatly separate children's distinct approaches to the series of problems. Here, we tested whether children produce similar distributions of response patterns on the FIM-An test, with criteria adjusted for the smaller number of FIM-An test trials (flexible: $\geq 83\%$, $\geq 63\%$; partly flexible: $\geq 58\%$, $\geq 37\%$; perseverative: $\leq 50\%$, $\leq 25\%$, $\leq 37\%$; indiscriminate: $\leq 50\%$, $\leq 25\%$, $\geq 50\%$). We predicted that all four patterns would occur in the FIM-An test, as in the FIM-Ob test. Our main question of interest was whether individual children would produce similar patterns in both tests.

RESULTS

Gender

There were no reliable gender differences in any measures in any test, so girls and boys were combined in all subsequent analyses.

FIM-An test

Overall, children were moderately sensitive to phrase cues: they averaged 54% cue-appropriate inferences ($SD = 24\%$), which is greater than chance (25%, or 33% if foil items are discounted). However, there are large between-cue differences. Children's proportions of cue-appropriate response averaged 90% in 'is a ...' trials, 39% in 'is in a ...' trials, and 34% in 'has a ...' trials ($SDs = 23\%$, 39%, 34%). Thus, children consistently mapped 'is a ...' to species, but did not consistently use the other cues. Cue-appropriate responses were not strongly correlated with age ($r = .25$, $p = .073$). The relation to age was somewhat cue-dependent, but not significantly so ('is a ...' $r = .01$; 'lives in/on a ...' $r = .26$, $p = .066$; 'has a ...' $r = .21$, $p = .135$). Cue-appropriate inferences were modestly correlated with raw PPVT scores ('is a ...' $r = .22$; 'lives in a ...' $r = .26$, $p = .066$; 'has a ...' $r = .20$; all cues, $r = .32$, $p = .022$). Neither Stroop Day-Night accuracy nor RT were significantly correlated with any measure of cue-appropriate responses.

Children might use cues differently if they precede the first, second, or third word for a stimulus set (see Figure 3). Proportion of cue-appropriate responses in each block are shown in Figure 4. There were significant cue effects in block 1 ($F(2,48) = 45.5$, $p < .001$, $\eta^2_{part} = .64$), in block 2 ($F(2,48) = 5.6$, $p = .006$, $\eta^2_{part} = .20$), and block 3 ($F(2,48) = 10.8$, $p < .001$, $\eta^2_{part} = .33$). Post-hoc Tamhane D_3 tests showed that in any block, children made more cue-appropriate responses to 'is a ...' than to 'is in/on a ...' or 'has a ...'. The latter two cues did not differ in any block.

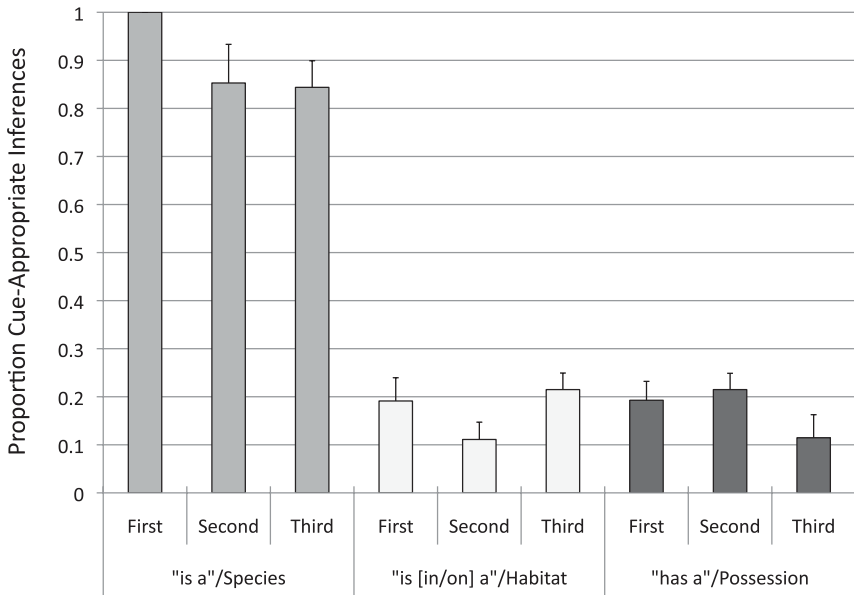


Fig. 4. Mean cue-appropriate responses in Experiment 1, for each block (i.e., relative to cue order). Each bar for a given cue represents a subset of approximately one-third of children. Error bars = *SE*.

These results suggest that in the FIM-An test, cue use was determined more by children's cue comprehension than by order-sensitive processes (e.g., flexibility). Overall cue-appropriate responses did not decline from the first to later blocks: means in blocks 1, 2, and 3 were, respectively, 54% ($SD = 43\%$), 54% (44%), and 55% (42%). The within-subjects effect was not significant ($\eta_{part}^2 < .01$), nor was the age covariate ($p = .067$, $\eta_{part}^2 = .07$), in a repeated-measures ANOVA. Notably, there was no Age \times Block interaction ($p = .875$, $\eta_{part}^2 < .01$), suggesting that cue use did not disproportionately decline (e.g., due to perseveration) in later blocks.

FIM-An flexibility

To evaluate parametric differences in flexibility (i.e., CORSWOPS), we considered only children who made at least 50% appropriate inferences in the first block (chance = 25% or 33%). It is very difficult to assess the flexibility of children who did not make cue-appropriate inferences even in the first block. The remaining thirty children were 61% of the entire sample; with a mean age of 4;1 (slightly older, not surprisingly, than the

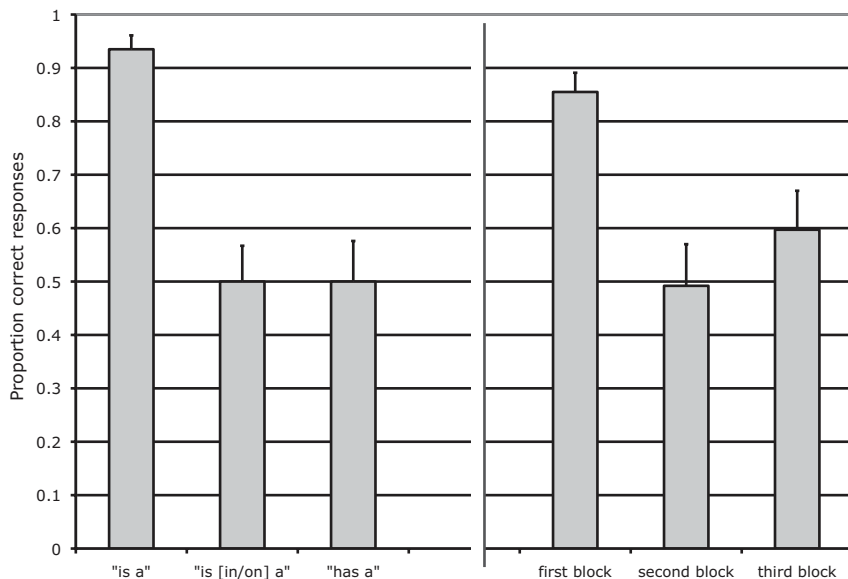


Fig. 5. Mean cue-appropriate responses (and *SE*) by the first-cue-using group (i.e., $\geq 50\%$ in block 1; $N=30$) for each cue, and in each block, in the FIM-An. Chance performance would be 0.25 (i.e., 25% correct).

entire group). They made more cue-appropriate inferences than expected (by the more-stringent 33% criterion) on every block: means = 87%, 50%, and 60% on blocks 1, 2, and 3, respectively (*SDs* = 19%, 44%, 41%). This group was also above chance (33%) for all three cues: mean for 'is a ...' = 92% (*SD* = 21%), 'in/on a ...' = 52% (38%), 'has a ...' = 52% (42%) (see Figure 5). Thus, this higher-performing subset of children tended to use all three cues to some extent, but used later cues less than the first cue. This suggests some effect of cognitive flexibility.

To further examine flexibility, we tested CORSWOPS ratios in an ANCOVA, with First-Cue between-subjects (see Figure 6), and age entered as a covariate. The First-Cue effect was not significant ($p = .123$, $\eta^2_{part} = .15$).⁴ The age covariate was not significant ($p = .152$, $\eta^2_{part} = .08$). The full model accounted for $R^2_{adjusted} = .11$. Thus, other factors account for differences in flexible use of word-meaning cues.

[4] Power is limited because relatively few children who had 'has a' cue in block 1 met the 50% criterion. However, the pattern of results is similar if we consider the entire sample of $N=51$.

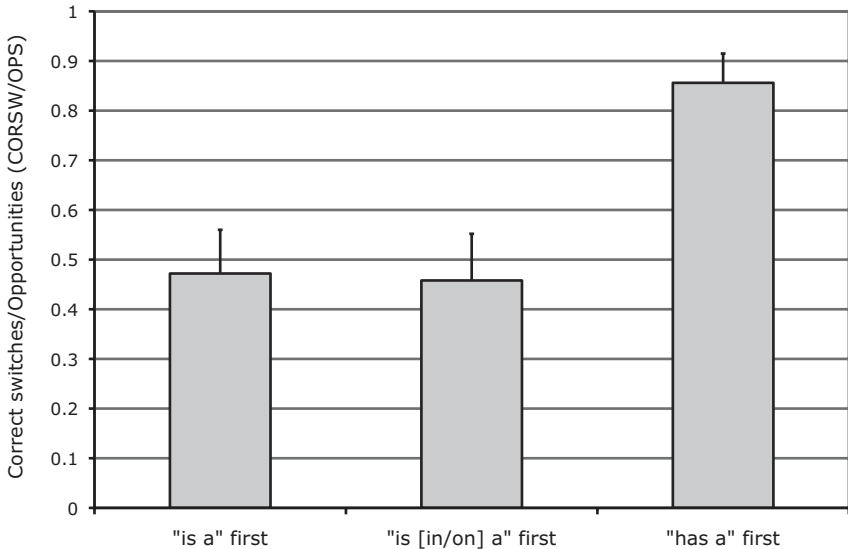


Fig. 6. Mean (and *SE*) proportion of CORSWOPS in the FIM-An by the first-cue-using group, divided by the first phrase cue ('is a', $n=19$; 'is [in/on] a', $n=8$; 'has a', $n=4$).

FIM-Ob test

As reported by Deák and Narasimham (2003), children made more cue-appropriate responses in the FIM-Ob test than expected by chance (mean = $\cdot60$, $SD = \cdot21$). They also showed smaller differences between the three cues, compared to the FIM-An test: means for 'is a'/shape, 'is made of'/material, and 'has a'/part were $\cdot48$ ($SD = \cdot35$), $\cdot69$ ($\cdot34$), and $\cdot64$ ($\cdot33$), respectively. Cue-appropriate responses were moderately correlated with age ($r = \cdot492$, $p < \cdot001$). This relation was somewhat cue-dependent ('is a ...': $r = \cdot273$, $p = \cdot053$; 'is made of ...': $r = \cdot456$, $p = \cdot001$; 'has a ...': $r = \cdot181$, $p = \cdot199$). Cue-appropriate inferences also were correlated with PPVT scores ($r = \cdot518$, $p < \cdot001$).

The thirty children who tended to use the first FIM-An test cue were also responsive to the FIM-Ob test cues, averaging $\cdot72$, $\cdot64$, and $\cdot61$ cue-appropriate inferences in blocks 1, 2, and 3, respectively ($SDs = \cdot32$, $\cdot33$, and $\cdot25$); all above chance (33%) at $p < \cdot001$. A multivariate test reveals that the decline in cue use across blocks was not reliable ($F < 1$). However, there was a significant age effect: younger children were less likely than older children to utilize later phrase cues.

TABLE 1. *Simple and partial correlations among FIM-An and FIM-Ob measures of flexibility, PPVT-R raw score, and Stroop accuracy and response time (Experiment 1, filtered sample)*

Correlation	FIM-An	FIM-Ob	PPVT-R	Stroop correct	Stroop RT
Age	.23	.35	.51***	.32	-.32
FIM-An CORSWOPS		.58** [.53**]	.32 [.21]	.15 [-.07]	-.18 [-.13]
FIM-Ob CORSWOPS			.46** [.28]	.27 [-.03]	-.32 [-.27]
PPVT-R standardized				.12 [.03]	-.40* [-.56**]
Stroop correct					-.31 [-.29]

NOTES: * $p < .05$; ** $p < .01$; *** $p < .005$. Coefficients in brackets are partial correlations, controlling for age.

FIM-Ob test CORSWOPS ratios averaged .57 ($SD = .29$), similar to the FIM-An test mean of .52. An ANCOVA comparing the different cue orders found no reliable order effect ($F < 1$).

Comparing FIM tests

We compared performance on the two FIM tests to determine whether they assess similar word-learning capacities. For a more conservative test, we focused on the children who used the first FIM-AN test cue. However, the findings reported here are the same for the entire sample.

Correlations were calculated between normalized CORSWOPS scores in both FIM tests, PPVT scores, and Stroop accuracy and RT. Simple correlations are shown in Table 1. Because age was positively related to all measures, it was partialled out. There was a significant partial correlation between CORSWOPS in the FIM-An and FIM-Ob tests ($r_{\text{partial}} = .530$, $p = .008$). Even with age and vocabulary partialled out, the correlation remained strong ($r_{\text{partial}} = .502$, $p = .015$). There was no relation between any measure of flexibility in either test, and either Stroop accuracy or RT. Receptive vocabulary (PPVT) was positively but non-significantly correlated with FIM flexibility and with Stroop RT.

Response patterns for individual children in both FIM tests are shown in Table 2. Fifteen children (50%) produced the same pattern on both tests – almost twice the expected number (7.7). This is a significant difference (χ^2 [df = 9; $N = 30$] = 20.4, $p = .016$). This suggests moderate within-subject consistency across the two tests.

DISCUSSION

Our primary question was whether children use phrase cues to constrain their inferences about word meanings when those cues and their implications

TABLE 2. *Cross-tabulation of children's response patterns in both FIM tests, Experiment 1. Children were selected based on using the first FIM-An cue (see text)*

FIM-An FIM-Ob	Flexible	Partly flexible	Indiscriminate	Perseverative	Totals
Flexible	5 [16.7%]	4 [13.3%]	0 [0]	0 [0]	9 [26.7%]
Partly flexible	2 [6.7%]	5 [16.7%]	1 [3.3%]	4 [13.3%]	12 [40.0%]
Indiscriminate	0 [0]	1 [3.3%]	2 [6.7]	3 [10.0%]	6 [16.7%]
Perseverative	0 [0]	0 [0]	0 [0]	3 [10.0%]	3 [10.0%]
Totals	7 [23.3%]	10 [33.3%]	3 [10%]	10 [33.3%]	30

change, and possibly interfere, across successive inferences. We also considered whether children's use of changing phrase cues was modulated by shifting to stronger versus weaker cues, and by a bias to focus on one property (e.g., animal species).

Results from the FIM-An test indicate that some three- and four-year-olds can use changing phrase cues to infer different meanings for different words, even if the cues are subtly different (e.g., 'is a ...' vs. 'has a ...'), and children are biased to choose one stimulus property (i.e., species). Children across the entire group made more cue-appropriate inferences than expected by chance. However, children varied considerably in cue-utilization in both FIM tests, and were not as a group above chance in mapping the 'is [in/on] a ...' or 'has a ...' cues to habitats and possessions in the FIM-An task. This variability in cue use was correlated with age, presumably because semantic knowledge continually increases, but the relation was modest. There also was a modest correlation with receptive vocabulary, even when age was partialled out. Still, both variables only accounted for a small proportion of variance.

When children who did not utilize the first cue were removed from the analysis, sizeable individual differences remained. Response patterns found in the FIM-Ob test – flexible, partly flexible, indiscriminate, and perseverative – were also seen in the FIM-An test. Individual children tended to produce similar response patterns across tests. More generally, measures of flexibility on the FIM-An and FIM-Ob tests were moderately to strongly correlated, even with age, vocabulary, and Stroop accuracy partialled out. This is noteworthy because the tests use different ontological kinds, properties, and stimulus types (objects vs. pictures). The correlation therefore suggests some abstract skill or sensitivity. One possibility is that children who do not comprehend the implications of the cues very clearly in one task will probably not comprehend them very clearly in the other. However, this is unlikely to explain all of the results, for at least two reasons: first, the PPVT (a measure of receptive language skill) only

modestly predicted accuracy in either FIM test; second, the CORSWOPS measures were correlated even with cue comprehension effects controlled.

Although there is some continuity of individual performance between the tests, there is also non-shared variance. As noted above, one factor is a difference in cue-property 'strength' between tests. In the FIM-Ob test, children as a group were above chance in mapping each cue to its implied property (Deák, 2000), whereas this was true of 'is a'/species mappings in the FIM-An test. This finding seems to reflect a species bias: children made three times as many same-species responses as habitat- or possession-based responses in the FIM-An test. The FIM-Ob test does not elicit such a strong property-specific bias. To the extent that the property itself caused the bias, we cannot say whether it is because species matches are generally compelling (Gelman, 2003), or because the same-species pictures were perceptually salient (i.e., in the foreground, and presenting many distinctive properties).

One reason why the FIM-An test might have been somewhat harder than the FIM-Ob test is that the cues in the former were more subtly different. 'is a ...', 'is [in/on] a ...', and 'has a ...' differed only by one morpheme, and children who are younger, or have less developed receptive language, might not have noticed or processed these differences. Experiment 2 tests this possibility by administering a revised FIM-An test with phrase cues that were more semantically specific and more lexically differentiated. The habitat cue was changed to 'lives [in/on] a ...', which uses a more specific verb phrase, and the possession cue was changed to 'holds a ...', which more clearly implies the grasped object. This change addresses an important concern: in Experiment 1, a substantial proportion of children had to be excluded from analyses of flexibility. This limited our ability to make inferences about the age differences. For this reason, it remains unclear whether the FIM-An test can validly assess flexible cue-based word learning in most children younger than four years. By using more distinctive phrase cues in Experiment 2, we can assess whether a revised version of the FIM-An test can measure word-learning flexibility in three-year-olds as well as four-year-olds.

EXPERIMENT 2

To determine whether the findings of Experiment 1 are reliable, and whether the FIM-An test can assess flexible cue use in three-year-olds as well as four-year-olds, a modified version was given to a new group of children. In the modified version, two of the phrase cues were made more distinct and specific. First, 'lives [in/on] a ...' replaced 'is in a ...'. Also 'holds a ...' replaced 'has a ...'. Changing the phrase cues allowed us to test the strength of children's same-species bias. Could children ignore this property, and the strong cue-to-property association? It is unclear whether children's bias to

perseverate on the species match will generalize to a modified cue set: to the extent that the other two cue–property associations are stronger, children might be more likely to attend to similarities other than the species match.

In addition, two new stimuli sets with different features were created, to make the FIM-An test more statistically sensitive, and to most powerfully test the generalizability of the results.

METHOD

Participants

Thirty-six children, including eighteen three-year-olds (10 girls; mean age = 3;8, range 3;3–3;11) and eighteen four-year-olds (6 girls; mean = 4;9, range 4;0–5;4) were recruited from preschools in San Diego, CA. Children were fluent in English and were primarily European-American and middle class.

Materials

Six sets of pictures were used. These were the same four as in Experiment 1, plus two new sets with the same structure. Species, habitats, and possessions were made as different as possible (e.g., species included quadruped, arboreal, aquatic, flying, and vaguely humanoid types). (To view or download the full stimulus sets, go to <http://www.cogsci.ucsd.edu/~deak/cdlab>.)

Procedure

Each child was tested individually in a quiet area of preschool. The FIM-An test was given as in Experiment 1, with the following changes: children completed three blocks of six trials, rather than four trials, for a total of eighteen trials with a unique novel word in each trial. Each block featured a different phrase cue: ‘is a ...’, ‘lives [in/on] a ...’, and ‘holds a ...’. Cue order was counterbalanced. Randomization and counterbalancing, and coding, were the same as in Experiment 1.

RESULTS

Preliminary *t*-tests revealed no significant or marginal gender differences in cue accuracy or CORSWOPS. Thus, boys and girls were combined in subsequent analyses.

The entire group averaged 58.9% total cue-appropriate choices ($SD = 24.4\%$), which is above chance ($t(35) = 8.2$, $p < .001$). Their CORSWOPS ratios averaged .44 ($SD = .39$). The sample was filtered, as in Experiment 1, to make the analyses more valid and conservative. Children with at least 50%

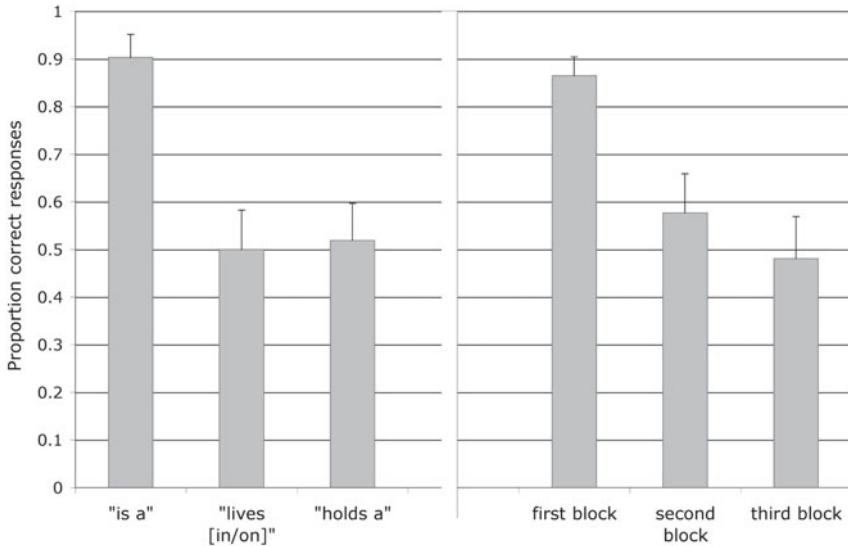


Fig. 7. Mean correct responses (and *SE*) to each phrase cue, and in each block, in the FIM-An in Experiment 2. Data include the filtered sample (i.e., $\geq 50\%$ in block 1), $N=27$.

first-cue choices made up 72.2% of the sample ($n=26$). Their average age was 4;3. They averaged 11.4 of 12 opportunities to switch, making CORSWOPS a valid measure of flexibility. The filtered sample performed slightly better than the entire group (mean = 63.9% cue-appropriate choices, $SD=25.6\%$; CORSWOPS mean = .52, $SD=.40$). Thus, filtering had only modest effects on the results.

The filtered sample's mean appropriate responses to each cue, and to each trial block, are shown in Figure 7. Children made more appropriate 'is a ...' responses (mean = 90.3%; $SD=24.6\%$) than 'lives in a ...' (50%; 42.2%) or 'holds a ...' (52%; 39.5%) ($F(3,22)=13.3$, $p<.001$). However, both of the latter were above chance ($t(25)=3.0$ and 3.5 , $p=.006$ and $.002$), unlike in Experiment 1. Three-year-olds made fewer appropriate inferences than four-year-olds based on the 'lives in a ...' cue (25.8% vs. 67.8%, $t(21.5)=2.8$, $p=.010$) and the 'holds a ...' cue (33.3% vs. 65.5%, $t(20.9)=2.2$, $p=.040$). This suggests that the more distinctive cues were moderately effective at elicited phrase-appropriate responses, but only from four-year-olds. However, three- and four-year-olds did not differ in species-based responses to 'is a ...' words (92.5% vs. 88.8%, $t<1$). This is consistent with the bias toward same-species responses seen in Experiment 1.

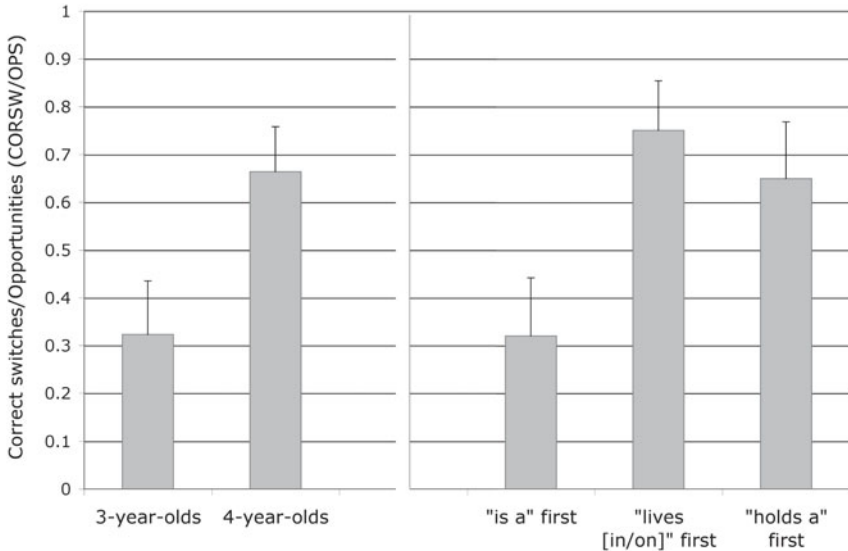


Fig. 8. Mean (*SE*) CORSWOPS in the FIM-An, Experiment 2. Left: three-year-olds ($n=11$) and four-year-olds ($n=17$) Right: groups by block 1 cue ('is a', $n=19$; 'is [in/on] a', $n=8$; 'holds a', $n=4$).

Cue-appropriate responses in the last two blocks were submitted to a 2 (age) \times 3 (First Cue) ANOVA. The age effect was not significant ($F(1,26)=1.5$, $p=.23$, $\eta_{part}^2=.07$), but the First Cue effect was significant ($F(2,26)=4.6$, $p=.023$, $\eta_{part}^2=.31$). Mean appropriate responses in the later blocks averaged 32.6% after the 'is a ...' cue, 76.3% after the 'lives [in/on] a ...' cue, and 65.6% after the 'holds a ...' cue ($SDs=41.7\%$, 23.8%, and 28.7%, respectively). The difference is because, in the latter cases, one of the two later cues was 'is a ...'. Post-hoc comparisons showed a significant difference between the 'is a ...' and 'lives [in/on] a ...' First-Cue groups ($p=.035$ by Bonferroni test); no other First-Cue groups were significantly different. The Age \times First-Cue interaction was not significant.

Flexibility (CORSWOPS) was examined in a 3 (First Cue) \times 2 (three vs. four years) ANOVA. Means for each age group and each First-Cue group are shown in Figure 8. The First-Cue effect was significant ($F(2,20)=3.9$, $p=.037$, $\eta_{part}^2=.28$). Dunnett's T_3 post-hoc tests showed a difference in flexibility between the groups that heard 'is a ...' first and 'lives in a ...' first ($p=.051$). The age effect was not significant ($F(1,20)=1.2$, $p=.288$, $\eta_{part}^2=.06$), though it was in the expected direction ($r=.213$). The Age \times First Cue interaction was not significant ($F(2,20)=1.1$). The complete model accounted for $R^2=.28$.

Children were classified as flexible, partly flexible, indiscriminate, or perseverative, based on the criteria for the FIM-Ob test in Experiment 1 (because there were now six FIM-An test sets). Overall, 31% of children were flexible (9% of three-year-olds; 53% of four-year-olds), 27% were partly flexible (27%, 27%), 8% were indiscriminate (18%, 0%), and 31% were perseverative (45%, 27%). The distributions for the three- and four-year-olds were not significantly different ($\chi^2[N=26, df=3]=7.1, p=.070$). Notably, 38% of children produced a non-canonical (partly flexible or indiscriminate) pattern. This supports previous findings that children are not 'just' flexible or perseverative (Deák, 2003). As in Experiment 1, every child who perseverated focused on species, even with the clearer habitat- and possession-implying cues.

DISCUSSION

Children's use of phrase cues generally replicated Experiment 1, with newly added stimuli sets and more specific, distinctive phrase cues. Over 60% of children who attended to the first cue switched a substantial number of responses to at least one of the later phrase cues. Also, seven out of twelve (58%) children in the entire sample who extended 'is a ...' words to species in the first block later adapted their inferences to at least one of the later cues. Those seven children were, on average, 53.5 months old. This suggests that by 4 1/2 years of age, many children can effectively subjugate interfering response biases or previous inferences in favor of local cues. In addition, 36% of three-year-olds (29% of all three-year-olds) produced flexible or partly flexible patterns, versus 80% of four-year-olds (72% of all four-year-olds). This indicates a robust age-related trend in children's responses to successive problems (i.e., inferences) about the same stimuli. Despite these findings, age differences were smaller than individual differences.

Although the modified phrase cues ('lives in a ...', 'holds a ...') were more distinctive and specific, many children still did not use them, particularly after making same-species responses. This indicates a bias to map novel words to species, at least in these stimuli.

GENERAL DISCUSSION

Cognitive flexibility is an integral aspect of language learning and language use (Deák, 2003; Jacques & Zelazo, 2005). The present study was designed to probe the breadth and consistency of three- and four-year-old children's flexible use of intra-sentential cues to infer novel noun meanings. The use of intra-sentential phrase cues was motivated by the fact that in everyday speech, meaningful linguistic elements – words and phrases – surround and constrain unfamiliar words. These local contextual elements are, we believe, among the

most critical cues to word meaning. However, these cues are variable, unpredictable, and changeable. To use them for word learning, children must select the right semantic cue at the right time, and draw the most relevant and correct implications. This entails cognitive flexibility. Flexibility is, however, resource-demanding even for older children and adults (Coulson & Kutas, 2001; Singer, Graesser & Trabasso, 1994). It would not be surprising if it were found to exceed the cognitive resources of young children.

Young children can in fact use simple cues to infer word meanings (Goodman *et al.*, 1998). However, they also make some curious errors. Some errors seem to privilege more distant information – earlier phrase cues or responses – over local cues. It has been unclear why young children make such errors, and how they become more flexible, particularly from two to six years of age. Two possible, non-mutually-exclusive causes are the accumulation of semantic knowledge, which permits the use of a wider range of phrase cues, and the maturation of cognitive flexibility processes, which support adaptation to changing cues. The present study extends previous work to understand the growth of flexibility in using phrase cues to learn words (Deák, 2000). For this we developed a new test (the FIM-An) to compare to an existing test (FIM-Ob).

The results showed considerable variability in three- and four-year-old children's responses to different cues, and in their flexibility across changing cues. There were prominent effects of cue difficulty, supporting the claim (Deák, 2003) that cue-to-property association strength is a critical factor in adaptive cue use. For example, in Experiment 1, the same cue, 'is a ...', was the weakest cue in the FIM-Ob test, where it implies an object-shape category, but the strongest cue in the FIM-An test, where it implies a species category. Also, cue strength differences appeared to be larger in the FIM-An test than in the FIM-Ob test. This seemed to contribute to an order effect: if 'is a ...' was the first cue in the FIM-An test, many children continued mapping later words onto species. Yet the 'is a'/species mapping did not cause intractable interference. Rather, children's use of cues in the FIM-An test was not affected by whether a cue occurred in the first or a later block (Figure 4). In other words, cue strength mattered regardless of whether the previous cue had been stronger or weaker. This is consistent with findings from a distinctly different study. Deák *et al.* (2004) found that three- and four-year-olds' flexibility in object sorting responses depended on the absolute difficulty of each sorting task rather than whether the task was first or second. It is also consistent with studies of adults' task switching, which find larger effects of task difficulty (e.g., 'compatible' vs. 'incompatible' trials) than direction of switching (i.e., easier-to-harder or vice versa) on response time (e.g., Yeung & Monsell, 2003). In fact, adults' RTs depend more on the strength of each cue than on whether or not the cue requires a response-set switch at all (Arrington, Logan & Schneider,

2007). This demonstrates the importance of cue strength. Chevalier and Blaye (2009) reported a similar finding in older children: transparent verbal task cues elicited faster and more accurate responses than less transparent or arbitrary cues, on both switch and 'stay' trials. Thus, for children and adults, cue strength is as important or more important than flexibility in determining how people adapt to a series of problems. The current results extend this generalization to young children's use of phrase cues for word learning.

Although the foregoing implies that children had great difficulties in the FIM-An test, it is worth noting that about one-third of children who heard 'is a ...' in the first block subsequently made cue-appropriate responses to 'lives in a ...' and 'holds a ...' cues (Experiment 2). Thus, a strong first cue does not necessarily shut down preschoolers' flexibility, though it seems to reveal large individual differences. These differences are being explored in a larger study.

In addition to phrase cue strength effects, we found between-test consistency in children's flexible cue use. The CORSWOPS ratios on the two FIM tests were correlated at $r = .50$, controlling for age and vocabulary, in children who were selected for their use of the first cue (as a more conservative test; the correlation was slightly higher in the whole sample). This correlation can be attributed to individual differences in flexibility, not baseline cue comprehension, for several reasons. First, the CORSWOPS measure controls some cue-specific effects, for example spurious cue-appropriate responses on later trials. The measure also controls for some baseline cue-order effects by considering the proportion of opportunities to switch rather than the absolute number. Second, the between-test correlation remained similar when first-block accuracy, and/or cue order effects, were partialled out or controlled (or, in more elaborate analyses available by request, by weighting each switch based on its expected difficulty). These measures control for residual effects of specific cue comprehension. Third, in non-parametric tests (Experiment 1), many children produced the same response pattern in both tests. These patterns reflect categorically different approaches to the task, which are not necessarily revealed by numbers of cue-appropriate inferences or switches. For example, perseverative and indiscriminate patterns might stem from different problems: perseveration from an over-reliance on prior cues, and indiscriminate responding from uncertainty about cue meanings. In any case, about half of children produced the same pattern on each test (though the FIM-An test was harder than the FIM-Ob test), and 73% were either flexible (partly or entirely) in both tests or inflexible (indiscriminate or perseverative) in both. Almost all of the remaining children were partly flexible in one test and inflexible in the other, which is what one would expect of children who are developing flexibility: they might adapt to changing cues that are easier, but fall back on an inflexible response when cues are harder.

The age differences were limited and measure-specific. They were not significant in parametric measures of flexibility in the FIM-An test. However, the distribution of different response patterns differed between three- and four-year-olds. This is consistent with previous studies of the FIM-Ob test that have shown (roughly) a reduction from three to four years in indiscriminate responding, then a reduction from four- to five-year-olds in perseveration, with a simultaneous increase in partial flexibility, and by five years a preponderance of fully flexible cue use (Deák, 2003).

The results also indicate test factors that go beyond cue semantics. Children's tendency to perseverate on species cannot just be due to cue strength: although children know that 'is a ...' can imply a biological kind (Gelman, 2003) with a generic label (Gelman *et al.*, 1998), 'is a ...' can equally well imply an object (e.g., a held object). Yet children preferentially mapped the words onto species, not the held objects. We believe this is because the creatures were the most salient features of the stimuli. They were large, distinctive, and prominently placed in the foreground. By contrast, the held objects were smaller, less detailed, and less prominent. Perceptual salience can influence how children map novel words onto properties (e.g., Jones *et al.*, 1992), and whether they will perseverate on a stimulus dimension (Brooks, Hanauer, Padowska & Rosman, 2003). This can explain the results. Moreover, it explains why the species bias was not eliminated by more specific and distinctive phrase cues (Experiment 2). That is, 'holds a ...' did not elicit more held-object inferences than 'has a ...' (Experiment 1) (55% vs. 57%), and 'lives in a ...' did not elicit more habitat inferences than 'is in a ...' (53% vs. 51%). If perceptual salience was overwhelming, changes in cue wording might not matter. (Of course, it is also possible that the revised cues were still too subtle.)

Given these considerations, we can recast scope errors – a discourse-level effect – as more multifactorial effects. Factors that potentiate these errors include children's cue comprehension and discourse knowledge, changes in cue strength, and stimulus salience or prior biases for certain properties.

Even if perseverative errors are caused by multiple factors, the functional consequence is a fundamental error in task pragmatics. That is, to adults the task should create a demand for different responses on different trials. Yet some children seem to act according to the opposite demand – an odd assumption that several distinct words refer to the same property, not different ones. Why is this? One possibility is that children perseverate as a fall-back strategy when they are unsure of a cue's meaning. If children believed that their previous response was correct, and the current question is too difficult, this would not be an unreasonable strategy (Deák, 2003). We cannot, however, test this hypothesis with the current data. This underscores the general difficulty of interpreting perseveration. Perseveration is typically ambiguous in the binary rule-switching tests that dominate the literature on

cognitive flexibility. This is because perseveration is the only alternative to a correct response. In the FIM paradigm, however, children choose between several responses, and adapt to several cues. Thus, perseveration is not inevitable, and is therefore a more informative behavior. For example, as we noted above, perseveration is most likely when a strong cue is given early, and one referent property is more salient. From these and other FIM results, we can also conclude that perseveration does not reflect an inability to inhibit prior responses (see Deák, 2003; Deák & Narasimham, 2003). Also, there was no correlation between any measure on the FIM and the Stroop Day–Night test, which supposedly measures children’s ability to inhibit verbal associations. Thus, the results add to converging evidence that inhibitory processes do not necessarily predict children’s perseveration (see also Cepeda *et al.*, 2001; Deák & Narasimham, 2003).

Some important questions remain. First, we cannot specify exactly how much the between-test correlation is due to cue comprehension, and how much to cognitive flexibility. This would require more extensive investigations of children’s use of many different verbal cues in different contexts. Second, and related to this, we do not know whether, and how much, stable individual differences in flexibility contribute to language learning and vocabulary growth in the long run. The finding of a modest correlation between flexibility and receptive vocabulary is suggestive, but not conclusive. That correlation hints at longitudinal stability, and it would be useful to investigate whether flexible use of verbal cues in preschool can predict the ability to infer word meanings from a written context during elementary school (Cain *et al.*, 2004; Cartwright, 2002). Even though we know that young children use different contextual cues to infer verb meanings (e.g., Behrend, Harris & Cartwright, 1995; Naigles, 1996), we do not know, for example, whether individual differences generalize to children’s acquisition of verbs or other kinds of words. Third, we do not know the breadth of the cognitive flexibility capacity measured in the FIM tests. It might be specific to verbal inference, or it might be a more general skill. This could be addressed by comparing individual children’s performance on the FIM tests with other tests of cognitive flexibility, as well as carefully selected tests of potentially related executive functions such as verbal working memory, generalized processing speed, and/or verbal inhibition. The present study only touches on these questions, but ongoing research is addressing these questions using that strategy.

CONCLUSIONS

The robust differences in children’s responses to different phrase cues indicate that cue/property mapping strength is a predominant factor in children’s flexible use of intra-sentential phrase cues to infer novel word

meanings. However, the robust correlation between the FIM-An test and the FIM-Ob test also suggests that individual children's cognitive flexibility was somewhat stable across task tests. This finding adds to prior evidence of the validity of the FIM-Ob test paradigm for revealing a particular kind of flexibility that might be important for word learning and vocabulary growth. The FIM-Ob test yields similar findings across testing sites and settings, geographic regions and dialects, and experimenters. Its results are unaffected by procedural details such as the use of preliminary practice trials, cue order, specific novel words, task length, or the delay between trials (Deák, 2000; Deák & Narasimham, 2003). The current data show that the CORSWOPS measure can be used with different stimuli or cues to reveal a moderately stable capacity for the flexible use of semantic cues. Our next task is to determine in more detail how this capacity operates across verbal and non-verbal tasks, with larger and more diverse groups of children.

REFERENCES

- Akhtar, N., Carpenter, M. & Tomasello, M. (1996). The role of discourse novelty in early word learning. *Child Development* **67**, 635–45.
- Arnold, J. E., Brown-Schmidt, S. & Trueswell, J. (2007). Children's use of gender and order-of-mention during pronoun comprehension. *Language and Cognitive Processes* **22**(4), 527–65.
- Arrington, C. M., Logan, G. D. & Schneider, D. W. (2007). Separating cue encoding from target processing in the explicit task-cuing procedure: Are there 'true' task switch effects? *Journal of Experimental Psychology: Learning, Memory, and Cognition* **33**(3), 484–502.
- Baldwin, D. A. (1989). Priorities in children's expectations about object label reference: form over color. *Child Development* **60**, 1291–306.
- Barlow, W. D. & Summers, I. (1979). *Barlow's guide to extraterrestrials*. New York: Workman.
- Beck, I. L., McKeown, M. G. & McCaslin, E. S. (1983). Vocabulary development: all contexts are not created equal. *Elementary School Journal* **83**(3), 177–81.
- Behrend, D. A., Harris, L. L. & Cartwright, K. B. (1995). Morphological cues to verb meaning: verb inflections and the initial mapping of verb meanings. *Journal of Child Language* **22**, 89–106.
- Brooks, P. J., Hanauer, J. B., Padowska, B. & Rosman, H. (2003). The role of selective attention in preschoolers' rule use in a novel dimensional card sort. *Cognitive Development* **117**, 1–21.
- Cain, K. & Oakhill, J. V. (1999). Inference making ability and its relation to comprehension failure in young children. *Reading and Writing* **11**(5), 489–503.
- Cain, K., Oakhill, J. V. & Lemmon, K. (2004). Individual differences in the inference of word meanings from context: the influence of reading comprehension, vocabulary knowledge, and memory capacity. *Journal of Educational Psychology* **96**(4), 671–81.
- Campbell, D. T. & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin* **56**(2), 81–105.
- Campbell, R. N. & Bowe, T. B. (1983). Text and context in early language comprehension. In M. Donaldson, R. Grieve & C. Pratt (eds), *Early childhood development and education*, 115–26. Oxford: Basil Blackwell.
- Cartwright, K. B. (2002). Cognitive development and reading: the relation of reading-specific multiple classification skill to reading comprehension in elementary school children. *Journal of Educational Psychology* **94**(1), 56–63.

- Cepeda, N. J., Kramer, A. F. & De Sather, J. (2001). Changes in executive control across the life span: examination of task-switching performance. *Developmental Psychology* **37**(5), 715–29.
- Chevalier, N. & Blaye, A. (2009). Setting goals to switch between tasks: effects of cue transparency on children's cognitive flexibility. *Developmental Psychology* **45**(3), 782–97.
- Corrigan, R., Halpern, E., Aviezer, O. & Goldblatt, A. (1981). The development of three spatial concepts: in, on, under. *International Journal of Behavioral Development* **4**(4), 403–19.
- Coulson, S. & Kutas, M. (2001). Getting it: human event-related brain response to jokes in good and poor comprehenders. *Neuroscience Letters* **316**(2), 71–4.
- Davidson, M. C., Amso, D., Anderson, L. C. & Diamond, A. (2006). Development of cognitive control and executive functions from 4 to 13 years: evidence from manipulations of memory, inhibition, and task switching. *Neuropsychologia* **44**(11), 2037–78.
- Deák, G. O. (2000). The growth of flexible problem solving: preschool children use changing verbal cues to infer multiple word meanings. *Journal of Cognition and Development* **1**(2), 157–91.
- Deák, G. O. (2003). The development of cognitive flexibility and language abilities. In R. Kail (ed.), *Advances in child development and behavior*, vol. 31, 271–327. San Diego: Academic Press.
- Deák, G. O. & Narasimham, G. (2003). Is perseveration caused by inhibition failure? Evidence from preschool children's inferences about word meanings. *Journal of Experimental Child Psychology* **86**(3), 194–222.
- Deák, G. O., Ray, S. D. & Brenneman, K. (2003). Children's perseverative appearance–reality errors are related to emerging language skills. *Child Development* **74**(3), 944–64.
- Deák, G. O., Ray, S. D. & Pick, A. D. (2004). Effects of age, reminders, and task difficulty on young children's rule-switching flexibility. *Cognitive Development* **19**, 385–400.
- DeMarie, D. (2001). A trip to the zoo: children's words and photographs. *Early Childhood Research & Practice* **3**(1), 1–26.
- Dickinson, D. K. (1988). Learning names for materials: factors constraining and limiting hypotheses about word meaning. *Cognitive Development* **3**, 15–35.
- Dunn, L. M. & Dunn, L. (1981). *PPVT-R Manual*. Circle Pines, MN: American Guidance Service.
- Ellefsen, M. R., Shapiro, L. R. & Chater, N. (2006). Asymmetrical switch costs in children. *Cognitive Development* **21**(2), 108–30.
- Ellis Weismer, S. & Hesketh, L. (1993). The influence of prosodic and gestural cues on novel word acquisition by children with specific language impairment. *Journal of Speech and Hearing Research* **36**, 1013–25.
- Friedman, O. & Neary, K. R. (2008). Determining who owns what: do children infer ownership from first possession? *Cognition* **107**(3), 829–49.
- Gelman, S. A. (2003). *The essential child: origins of essentialism in everyday thought*. Oxford: Oxford University Press.
- Gelman, S. A., Coley, J. D., Rosengren, K. S., Hartman, E., Pappas, A. & Keil, F. C. (1998). Beyond labeling: the role of maternal input in the acquisition of richly structured categories. *Monographs of the Society for Research in Child Development* **63**(1), 1–158.
- Gelman, S. A. & Heyman, G. D. (1999). Carrot-eaters and creature-believers: the effects of lexicalization on children's inferences about social categories. *Psychological Science* **10**(6), 489–93.
- Gerstadt, C. L., Hong, Y. J. & Diamond, A. (1994). The relationship between cognition and action: performance of children 3¹/₂–7 years old on a Stroop-like day–night test. *Cognition* **53**, 129–53.
- Goodman, J. C., McDonough, L. & Brown, N. B. (1998). The role of semantic context and memory in the acquisition of novel nouns. *Child Development* **69**(5), 1330–44.
- Gray, S. (2004). Word learning by preschoolers with specific language impairment: predictors and poor learners. *Journal of Speech, Language, and Hearing Research* **47**(5), 1117–32.

- Grieve, R., Hoogenraad, R. & Murray, D. (1977). On the young child's use of lexis and syntax in understanding locative instructions. *Cognition* **5**(3), 235–50.
- Hay, D. F. (2006). Yours and mine: toddlers' talk about possessions with familiar peers. *British Journal of Developmental Psychology* **24**(1), 39–52.
- Huxham, M., Welsh, A., Berry, A. & Templeton, S. (2006). Factors influencing primary school children's knowledge of wildlife. *Journal of Biological Education* **41**(1), 9–12.
- Jacques, S. & Zelazo, P. D. (2005). On the possible roots of cognitive flexibility. In B. Homer & C. Tamis-LeMonda (eds), *The development of social cognition and communication*, 53–81. Mahwah, NJ: Lawrence Erlbaum.
- Johnson, K. E. & Mervis, C. B. (1994). Microgenetic analysis of first steps in children's acquisition of expertise on shorebirds. *Developmental Psychology* **30**(3), 418–35.
- Jones, S. S., Smith, L. B. & Landau, B. (1991). Object properties and knowledge in early lexical learning. *Child Development* **62**(3), 499–516.
- Kalish, C. W. & Gelman, S. A. (1992). On wooden pillows: multiple classification and children's category-based inductions. *Child Development* **63**(6), 1536–57.
- Landau, B., Smith, L. & Jones, S. (1998). Object shape, object function, and object name. *Journal of Memory and Language* **38**(1), 1–27.
- Levy, B. A., Abello, B. & Lysynchuk, L. (1997). Transfer from word training to reading in context: gains in reading fluency and comprehension. *Learning Disability Quarterly* **20**(3), 173–88.
- Nagy, W. E., Anderson, R. C. & Herman, P. A. (1987). Learning word meanings from context during normal reading. *American Educational Research Journal* **24**(2), 237–70.
- Naigles, L. (1996). The use of multiple frames in verb learning via syntactic bootstrapping. *Cognition* **58**, 221–51.
- Samuelson, L. K. & Smith, L. B. (1999). Early noun vocabularies: do ontology, category structure and syntax correspond? *Cognition* **73**(1), 1–33.
- Saylor, M. M., Sabbagh, M. A. & Baldwin, D. A. (2002). Children use whole-part juxtaposition as a pragmatic cue to word meaning. *Developmental Psychology* **38**(6), 993–1003.
- Singer, M., Graesser, A. C. & Trabasso, T. (1994). Minimal or global inference during reading. *Journal of Memory and Language* **33**(4), 421–41.
- Smith, L. & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition* **106**(3), 1558–68.
- Storkel, H. L. (2001). Learning new words: phonotactic probability in language development. *Journal of Speech, Language, and Hearing Research* **44**(6), 1321–37.
- Strommen, E. (1995). Lions and tigers and bears, oh my! Children's conceptions of forests and their inhabitants. *Journal of Research in Science Teaching* **32**(7), 683–98.
- Thorndyke, P. W. (1976). The role of inferences in discourse comprehension. *Journal of Verbal Learning and Verbal Behavior* **15**(4), 437–46.
- Tomkins, S. & Tunnicliffe, S. D. (2007). Nature tables: stimulating children's interest in natural objects. *Journal of Biological Education* **41**(4), 150–55.
- Umstead, R. S. & Leonard, L. B. (1983). Children's resolution of pronominal reference in text. *First Language* **4**, 73–84.
- Yeung, N. & Monsell, S. (2003). Switching between tasks of unequal familiarity: the role of stimulus-attribute and response-set selection. *Journal of Experimental Psychology: Human Perception and Performance* **29**, 455–69.
- Yu, C. & Ballard, D. H. (2007). A unified model of early word learning: integrating statistical and social cues. *Neurocomputing* **70**(13–15), 2149–65.
- Zelazo, P. D. (2006). The Dimensional Change Card Sort (DCCS): a method of assessing executive function in children. *Nature Protocols* **1**(1), 297–301.
- Zelazo, P. D., Reznick, J. S. & Piñon, D. E. (1995). Response control and the execution of verbal rules. *Developmental Psychology* **31**(3), 508–17.