Cognition 127 (2013) 331–344

Contents lists available at SciVerse ScienceDirect

Cognition

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

Interaction between language and vision: It's momentary, abstract, and it develops

Banchiamlack Dessalegn[a,*], Barbara Landau[b]

[a] University of Chicago, Department of Psychology, 5848 S. University Avenue, Chicago, IL 60637, USA
[b] Johns Hopkins University, Department of Cognitive Science, 3400 North Charles Street, Baltimore, Maryland 21218, USA

ARTICLE INFO

Article history:
Received 2 November 2011
Revised 7 February 2013
Accepted 7 February 2013
Available online 29 March 2013

Keywords:
Whorf
Language
Development
Visual feature binding
Vision

ABSTRACT

In this paper, we present a case study that explores the nature and development of the mechanisms by which language interacts with and influences our ability to represent and retain information from one of our most important non-linguistic systems – vision. In previous work (Dessalegn & Landau, 2008), we showed that 4 year-olds remembered conjunctions of visual features better when the visual target was accompanied by a sentence containing an asymmetric spatial predicate (e.g., *the yellow is to the left of the black*) but not when the visual target was accompanied by a sentence containing a novel noun (e.g., *look at the dax*) or a symmetric spatial predicate (e.g., *the yellow is touching the black*). In this paper, we extend these findings. In three experiments, 3, 4 and 6 year-olds were shown square blocks split in half by color vertically, horizontally or diagonally (e.g., yellow-left, black-right) and were asked to perform a delayed-matching task. We found that sentences containing spatial asymmetric predicates (e.g., *the yellow is to the left of the black*) and non-spatial asymmetric predicates (e.g., *the yellow is prettier than the black*) helped 4 year-olds, although not to the same extent. By contrast, 3 year-olds did not benefit from different linguistic instructions at all while 6 year-olds performed at ceiling in the task with or without the relevant sentences. Our findings suggest by age 4, the effects of language on non-linguistic tasks depend on highly abstract representations of the linguistic instructions and are momentary, seen only in the context of the task. We further speculate that language becomes more automatically engaged in nonlinguistic tasks over development.

1. Introduction

In this paper we explore the interaction between two fundamental systems of human representation – language and vision. Although many cognitive scientists naturally assume that language and vision are independent, modular systems each with their own representational primitives and operations, most also acknowledge that the systems interact, supporting the uniquely human capacity to talk about what we see (Jackendoff, 1987; MacNamara, 1978). The broader question of how language interacts with vision and other cognitive systems has, for many decades, motivated scientists to ask about the role of language in influencing attention (Egeth & Smith, 1967; Gleitman, January, Nappa, & Trueswell, 2007; Lupyan, 2008; Lupyan & Spivey, 2010; Papafragou, Hulbert, & Trueswell, 2008; Smith, Jones, & Landau, 1996; Spivey, Tyler, Eberhard, & Tanenhaus, 2001), in sustaining conceptual categories and creating new ones (Roberson & Davidoff, 2000; Roberson, Davies, & Davidoff, 2000; Yoshida & Smith, 2003), and most radically, in creating new systems of representation (Carey, 2009; Hermer-Vazquez, Spelke, & Katsnelson, 1999). Implicit in all of these studies is the drive to understand the mechanisms by which language interacts with non-linguistic representations, and what the end result of the interactions is. Does language cause us to know the world differently? Does it allow us to represent categories and concepts in new ways that would be impossible without language? Does it leave intact the non-linguistic

* Corresponding author. Tel.: +251 920 848627; fax: +251 116 631722.
E-mail address: banchiamlack@gmail.com (B. Dessalegn).

0010-0277/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.cognition.2013.02.003

representational systems with which it has interacted, or does it change them?

Although these questions have been vigorously pursued in the recent literature on language and thought interactions, we still have quite limited understanding of the detailed mechanisms of these interactions. Perhaps this is not surprising: The causal chains running between language and thought are likely to be complex, varying with the particular claim being tested, the type of non-linguistic task one uses, the aspect of language that is under consideration, and whether the interactions are being studied in children or adults. In the end, there may be no "one size fits all" theory of the relationship between language and thought. But understanding anything about this relationship will require detailed study of the precise mechanisms by which language might interact with other systems of representation. Our goal in this paper is to provide a case study that demonstrates the nature and development of the mechanisms by which language might interact with and influence our ability to represent and retain information from one of our most important non-linguistic systems – vision. The particular domain of inquiry – representing and remembering the joint properties of color and location – lends itself to careful manipulation of the relative contributions of language and vision, enabling us to examine the conditions under which language does make a contribution, what aspects of language are operative in this contribution, and the developmental course by which language comes to play a prominent role in retaining visual elements.

To accomplish this goal, we frame our study in terms of several questions that pervade the literature on language–cognition interactions. Does language change the content and structure of the nonlinguistic representation with which it interacts? Is the effect of language momentary, seen only within the confines of a particular task, or must it always reflect long-term, permanent effects? What properties of language – syntax and/or semantics, – are relevant and responsible for the observed effect of language on cognition? Finally, are there developmental changes in whether (and the degree to which) language interacts with vision?

These questions are not new, and different answers to them have been offered in both the early and more recent literature. Indeed, these questions were at the center of Whorf's program, in which he argued that the language one speaks has a permanent restructuring effect on nonlinguistic thought. Under this hypothesis, the habitual use of language results in a parsing of the world that is consistent with one's language even when language is not being used. Implicit in this claim are answers to at least three of the questions we raised above: In Whorf's view, language changes the content of non-linguistic representations by "reshaping"; it has long-term effects that are difficult if not impossible to "undo"; and its effects are due to differences at all levels of language, including syntax and morphology.[1] Although Whorf did not explicitly consider whether the mechanisms of change were principally operative in children, it is easy to infer that they must operate as children learn language (and thought is thereby "reshaped"), but the same mechanisms may not be necessary in adults, whose thought has already been shaped by language. Of course, the exact details of any of these mechanisms were never completely clear from Whorf's theorizing, perhaps accounting for the ambiguity in whether recent findings are "Whorfian" or not.

1.1. Does language change thought?

More recently, the question of whether language changes thought has been addressed by two classes of proposals. In one, language is thought to induce a permanent restructuring of nonlinguistic thought including the domains of space and color (e.g., Davidoff, 2004; Davidoff, Davies, & Roberson, 1999; Levinson, Kita, Haun, & Rasch, 2002; Pederson et al., 1998; Roberson & Davidoff, 2000; Roberson et al., 2000). However, others have argued that the observed effects are due to methodological particulars, problems of interpretation of the data, and/or implicit use of language for so-called "non-linguistic" tasks (e.g., Li & Gleitman, 2002; Munnich & Landau, 2003). The second class of proposals posits qualitative change due to the presence of language without altering the non-linguistic representational base. These claims have been made in the context of number and space. Human understanding of number has been argued to undergo a radical transformation, bootstrapped through the number vocabulary (Carey, 2009); similarly, understanding of space has been argued to undergo radical restructuring as language enables the creation of new kinds of representations that combine existing but modularly separate cognitive domains (e.g., Hermer-Vazquez et al., 1999; Shusterman, Lee, & Spelke, 2011).

1.2. Are any effects of language permanent, or might they be momentary?

The question of whether language alters cognition permanently or has shorter, more time-bound effects has received relatively less attention, most likely because implicit in claims for restructuring is the idea that these changes are permanent (e.g., Davidoff, 2004; Davidoff et al., 1999). However, much evidence shows that some effects of language are quite short-lived (e.g., Dessalegn & Landau, 2008; Frank, Everett, Fedorenko, & Gibson, 2008; Frank, Fedorenko, Lai, Saxe, & Gibson, 2012; Landau, Dessalegn, & Goldberg, 2011; Li & Gleitman, 2002; Papafragou et al., 2008). For example Papafragou et al. showed a short-lived, task-specific effect of language. In their task, English and Greek speakers watched videos of motion events and were asked to carry out a linguistic or non-linguistic task. When participants engaged in language task, their pattern of eye movements changed depending on whether they spoke Greek or English, with English speakers showing more eye movements to the manner of motion, but Greek speakers showing more attention to the path. When people engaged in a non-linguistic matching task, however, they fixated the same aspects of the events regardless of their native language. Similarly, Li & Gleitman

[1] For example, one of Whorf's most famous claims concerned the role that the linguistic encoding of tense and aspect played in the cognitive representation of *time* among the Hopi. He claimed that critical aspects of time were not clearly represented in their language, and that this would naturally result in an absence of a concept of time in the culture (Whorf, 1956).

showed that English speakers' use of relative (e.g., the toy is to the *right* of the man) vs. absolute (e.g., toy is to the *east* of the man) frames of reference in a spatial task varied depending on the environment of testing, with both frames being used even within the same testing session – showing evidence of a flexible, task-dependent use of language. Finally, previous results from our lab have shown short-lived task-bound effects of language on vision (Dessalegn & Landau, 2008).

Specifically, in a series of studies we asked whether children's failure to encode and maintain conjunctions of visual features (a difficulty children share with adults) can be improved using language. A classic example of failure to bind was reported by Treisman and Schmidt (1982): If people are presented a very brief display containing a red O adjacent to a green L, they may mistakenly report seeing either a red L or a green O. This phenomenon is called illusory conjunction and was thought to reflect failure to bind color and shape (Treisman & Schmidt, 1982). Theories of visual attention, in particular the Feature Integration Theory (Treisman, 1977; Treisman & Gelade, 1980), have suggested that binding of visual features requires active allocation of focused attention at the location of the target object, with the location serving as the glue that binds the features together.

Our reasoning was that, if conjunctions of color and location were difficult for the visual system to represent and/or store in memory, then children might find it difficult to remember the exact color/location combination they saw, and hence they might make errors when asked to match the previously viewed sample to test items. Indeed, 4 year-olds found it quite difficult to remember color-location conjunctions over a delay as brief as 1 s. In the task, children were shown a block split in half by color, told to "look at it", saw it disappear, and 1 s later, were asked to match it to one of three test items including a replica of the block, its reflection, and a block with a different geometric split ("Other"; see Fig. 1). For these stimuli, distinguishing the target from its reflection requires correctly representing the relative spatial locations between the different colored parts of the block (e.g. top/bottom or left/right). Distinguishing the target from the different geometrically split block requires only representing the two different geometric splits (of target and "other"), but not the location of particular colors.

In this task, children matched correctly only about 65% of the time; their error patterns were predominantly reversals of color and location (i.e., high proportion of choosing the reflection), showing that they had specific difficulties retaining the combination of color and location of the block they had viewed. However, when we instructed them that "the red is on the left of the green", 4 year-olds improved substantially, now correctly matching on 80% of trials.

This instructing sentence explicitly expressed the spatial relationship between the two colored parts. In several additional experiments, we showed that the enhancing effects were quite specific to linguistic instruction of this particular form. For example, in one condition, children viewed the red section of the block flashing on and off for several seconds; in another condition, the red section repeatedly grew and shrunk; and in a third, children were

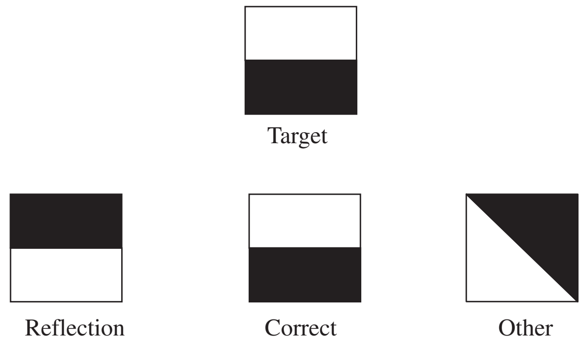

Fig. 1. Task design. The basic task design used in the current work and Dessalegn and Landau (2008). The target is presented at the top-center of a computer screen. The child views the target for unrestricted time then the target disappears followed by a 1 s delay followed by the appearance of the three test items at the bottom portion of the screen. The test items include a target replica (i.e., Correct response), a Reflection of the target, and second distracter (Other).

asked to "point to" the red part of the block before seeing the test items. If the enhancing effects we observed with "left/right" were general attentional effects, then these visual-spatial attentional manipulations should have also resulted in improvements beyond baseline. But performance in these conditions was no different from baseline, suggesting that the facilitating effects of language were not due to a general attentional mechanism. Furthermore, not all linguistic conditions were effective. Although describing the targets using "red on the left/ right" improved performance, children showed no improvement past baseline when the targets were labeled with novel whole object noun (e.g., "See this? This is a dax") or with spatial relational terms that were non-directional (e.g., "See this? The red is *touching/next to/connected to/up against* the green"). Finally, we found that the enhancement observed when children were instructed with "left/right" was not correlated with children's long-term knowledge of the spatial meanings of the terms. Specifically, in a post-test, we found that performance in the matching task was not correlated with whether children could correctly produce and comprehend terms "left" and "right. Whether they were correct or not in this post-test, they still benefitted from the linguistic description provided during the matching task, suggesting that the effect of linguistic instruction was short lived.

1.3. Are effects of language restricted to the lexicon, or are there effects of morphology and syntax?

Little direct research has focused on which specific properties of language are operative in causing effects on nonlinguistic representation, but there are arguments for effects at both the lexical and morphological levels. In general studies of color terms have focused exclusively (and naturally) on the lexical level, with comparisons across languages having different numbers of "basic terms" that divide the color space differently (e.g., Davidoff et al., 1999; Roberson et al., 2000; Rosch Heider & Olivier, 1972). Studies of spatial language have focused on specific lexical items (e.g., "in" or "on") or on phrases that encode different types of reference systems, for example, "to my right" vs. "east of me" (e.g., Levinson et al., 2002; Li & Gleitman, 2002), but have not explored the degree to which particular properties of phrasal syntax contribute to any effects. Studies examining object and substance biases have focused on the morpho-syntactic expressions that encode the distinction between count and mass nouns in English but not other languages such as Japanese (Yoshida & Smith, 2003) or Mayan (Lucy & Gaskins, 2003), but there is reason to believe that results showing effects on non-linguistic tasks may reflect statistical regularities in the nouns of the language (Barner, Inagaki, & Li, 2009; Gleitman & Papafragou, in press). In general, our understanding of whether lexical and/or syntactic properties contribute to effects of language on non-linguistic tasks is quite limited.

1.4. Is there developmental change in the use of language to modulate non-linguistic cognition?

With regard to changes over development, it has generally been assumed that language comes to play an increasingly strong role in modulating non-linguistic cognition. For example, memory researchers have long argued for a developmentally emerging increased influence of language on non-linguistic cognition (e.g., Flavell, 1970; Flavell, Beach, & Chinsky, 1966; see also Baddeley, Gathercole, & Papagno, 1998; Hagen & Kingsley, 1968; Reese, 1975). Reese (1975) found that children between the ages of 2;8 and 4;7 failed to produce useful verbal descriptions of visually presented spatial relationships to help them remember the pictures, whereas 5–6 year-olds automatically did so. Loewenstein and Gentner (2005) have found that spatial language (e.g., middle, top, bottom) helps 3 year-olds and 4 year-olds complete an easy spatial mapping task (but not a more difficult task) while language only helps 4 and 5 year-olds in the more difficult task.

The developmental period from age 3 through 6 thus appears to be a time of rapid development in the use of language to support better, more targeted encoding of the visual world, which could then aid in variety of tasks. By the time of adulthood, people appear to automatically encode visual stimuli using language (Stefurak & Boynton, 1986), raising the possibility that the developmental timeline supporting language–vision interactions might be quite protracted. Using language as a "tool" for recoding the non-linguistic world has been suggested to be a major developmental achievement (Gentner & Goldin-Meadow, 2003); but again, we still know little about the fine-grained nature of the developmental changes that take place as the child comes to carry out this "automatic" recoding.

1.5. The present studies

The findings from Dessalegn and Landau (2008) tell us several things about the role of language with respect to encoding visual stimuli. When 4 year-olds were instructed with the general comment "Look at this", and then asked to remember and match the target, they frequently erred by confusing the combinations of color and location. In contrast, when told "The red is on the left", their performance improved markedly, suggesting that language can have a powerful effect on 4 year-olds' encoding and/or memory for a visual stimulus that is otherwise quite fragile. The findings also tell us that the effects can be momentary, taking place only during the context of the task, but likely gone in the minutes after the task. Finally, the findings tell us that the effects appear to be remarkably specific to the particular linguistic instruction that is used. Neither non-linguistic attentional manipulations nor other expressions that are highly similar (e.g. "The red is touching the green") had the same effect on 4 year-olds' ability to match blocks to previously seen targets defined by their color and location combinations. This last finding raises questions about the precise nature of the linguistic representation that is facilitating the match, as well as the developmental time frame within which these effects can occur.

In the following experiments, we pursue these questions. The first concerns the nature of the linguistic representation that underlies our effect. Although we found that the effects are specific to directional terms such as "left" and "right" (and do not occur for predicates such as "touching" or "next to"), we do not know whether these expressions help the child because "right" and "left" encode the correct spatial relationship between the two parts, or whether they do so simply because they create an abstract representation that specifies some asymmetry between the two parts. Recall that the stimuli shown to children were *symmetric* – the targets were squares with a red part and a green part of the same size and shape and no inherent asymmetry (or figure-ground). Recall also that the only linguistic instruction to improve performance used *asymmetric* predicates such as "right/left/top/bottom".

These predicates define *nonreversible* and *asymmetric* relationships. More precisely, the label "X is to the left of Y" has three properties: (1) it defines a relationship over X & Y; (2) this relationship is asymmetric in that X and Y have different psychological status in the sentence (X is the figure and Y is the ground); and (3) the relationship is *nonreversible* such that if "X is to the left of Y" is true, then "Y is to the left of X" must be false. Similar sentences that were used with neutral terms (X is touching/next to/ connected to/ up against Y) engage verbs that can encode asymmetric relationships (in that they establish a figure and ground) but the relationship is reversible (if "X is touching Y" then "Y is touching X is also true").[2] Depending on the syntax that is used, the expression can define a relationship over X & Y which is either *asymmetric* or *symmetric.* The expression "X is touching Y" defines an *asymmetric* relationship over X & Y; that is, X and Y have different status in the sentence (X is the figure, Y is the ground). The expression "X and Y touch" also define a relationship over X & Y, but this is *symmetric*; that is, X and Y have the same status in the sentence (there is no clear figure or ground).

[2] We leave out the meaning of *touch* that is active, as in one person touching another with their hand. We refer to these terms as symmetrical because they *can* encode symmetrical and reversible relationships; this is a consequence of their core meaning. Unlike terms such as *left*/*right*, we can say "X is touching Y" or "X and Y touch", and in both cases, the sentences entail that both X touches Y and Y touches X. Terms like *left*/*right* cannot be used in the reversible sense; that is, we cannot say "X and Y are to the left" meaning "to the left of each other". For a complete discussion of this issue, see Gleitman, Gleitman, Miller, & Ostrin, 1996.

Crucially, however, neutral predicates such as *touch* (unlike the asymmetric and directional predicates "left/right"), define a *reversible* relationship such that the sentence "X is touching Y" entails "Y is touching X" (for more on asymmetric predicates see Gleitman et al., 1996; Miller, 1998).

Thus our previous work showed that given *symmetric* visual stimuli (where there is no inherent/natural asymmetry among the parts), only sentences containing the *asymmetric and non-reversible predicates* "top/bottom/left/right" led to improvements in children's memory for color-location information. If it is the spatial meaning of the predicates (the "left-ness") that is crucial to the enhancing effect, then we would not expect the effect to generalize to other terms that are non-spatial, even if they are asymmetric and non-reversible. In contrast, if the enhancing effects are due to the more abstract asymmetric character of the predicates, then a rather counter-intuitive prediction follows: Predicates that can establish a non-reversible asymmetry between the parts of the target – *even ones which do not convey spatial information* – should improve children's performance. For example, a similar performance benefit should be observed if we describe the targets using sentences with non-spatial asymmetric predicates, e.g. "the red is *prettier/lighter/brighter/nicer than* the green".

In Experiment 1, we test this prediction by asking whether sentences containing non-spatial asymmetric and non-reversible predicates (e.g., "the yellow is *prettier* than the black") facilitate 4 year-olds' performance. In Experiments 2 and 3, we examine the developmental course of the effects, extending our studies to younger children (Experiment 2) and to older children (Experiment 3). The results of the three experiments inform us about the nature of the linguistic representation underlying the enhancement in children's ability to represent and remember color/location conjunctions. They also shed light on how these representations operate during the developmental time period from 3 to 6, possibly a time of rapid developmental change in the use of language to support representation and memory.

2. Experiment 1A

In this experiment, we ask whether sentences containing non-spatial asymmetric and non-reversible predicates (e.g., "the yellow *is prettier than* the black") facilitate performance in representing and remembering color/ location conjunctions among 4 year-old children.

2.1. Participants

Thirty-six 4 year-olds (*M*age = 4;4, Range = 4;0–4;9) were randomly assigned to a No Label ($n = 12$), Directional label ($n = 12$), or Prettier label ($n = 12$) condition.

2.2. Stimuli, design, procedures

The stimuli, design and procedures were the same as in Dessalegn and Landau (2008), except for the variation in instruction (condition). Each of 6 possible targets was a square, 2.5 cm × 2.5 cm, split in two equal halves colored black or yellow, with either a vertical split (left/right partitioning), horizontal split (top/bottom partitioning) or a diagonal split (Fig. 2). Each trial was designed as shown in Fig. 1: The target was shown at the top of the computer screen, and the test stimuli (shown at the bottom) included a replica of the target (Correct) and two distracters (Reflection and Other distracters). The Reflection distracter had the same internal split as the target, but the color pattern was a reflection of the target. For example, if the target had a vertical split with yellow on the left side and black on the right, the Reflection distracter was a vertically split square with yellow on the right and black on the left. The Other distracter had an internal split that was different from the target. Each of the six targets was presented four times, for a total of 24 trials.

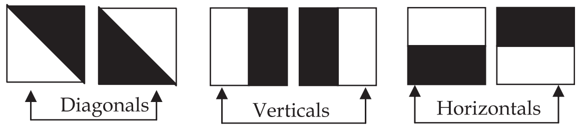

Fig. 2. Target blocks for current work. The figure shows a black and white reproduction of the 6 yellow-black blocks used in Experiment 1 and all subsequent experiments.

The three instructional conditions were No label, Directional label and Prettier label. Thus, the experiment was a 3-Label Conditions (No Label, Directional, Prettier) × 3-Target Type (Vertical, Horizontal, Diagonal) mixed design with label type as between-subjects variables and target type as a within-subjects variable.

In the No Label condition, participants were shown the target and instructed, "Look at this! Look at it carefully". The target then disappeared, followed by a 1-s delay, and then three test items appeared. In the Directional label condition, the relative location of the yellow part of the target (to the black) was labeled using directional terms, e.g. 'Look! The yellow is on the left/right/top/bottom of the black'. For the Vertical targets (split left/right by color), the position of the yellow part was labeled using 'on the left' or 'on the right'. For the Horizontal targets (split top/bottom by color) the yellow part was labeled using 'on the top' or 'on the bottom'. The yellow part of the Diagonal targets was labeled using 'on the left' and 'on the right'.[3] In the Prettier label condition, four non-spatial asymmetric predicates were used to replace the four directional predicates from the Directional label condition. The mapping was between the predicates (e.g., instead of *left*, *prettier* was used; instead of *right*, *lighter* was used). Participants were told: 'Look! The yellow is prettier/nicer/brighter/lighter than the black'. Instructions for the Directional and Prettier conditions were delivered as follows: The experimenter said "The yellow is…", and then clicked on the target, which triggered a recorded audio file that played the appropriate phrase labeling either the position of the yel-

[3] In a previous set of experiments, we had used either top/bottom or left/right to label the Diagonal targets but found no difference between these conditions, so we decided to use just left/right for these targets. For example in Fig. 2, the first diagonal block was labeled with "The yellow is on the left of the black" and the other diagonal block was labeled with "The yellow is on the right of the black" (note that Fig. 2 is a black and white reproduction of the yellow-black blocks).

low part (i.e., 'to the left of the black', 'to the right of the black', 'on the top of the black', or 'on the bottom of the black') or an attribute of the yellow part (i.e., 'prettier than the black' 'nicer than the black' 'brighter than the black' or 'lighter than the black'). In each of the three conditions, test item presentation was accompanied by the experimenter saying "Which one of these looks *exactly the same* as the one you just saw?" Test items stayed on the screen until participants responded.

Stimuli were presented on a Dell Inspiron 5100 laptop on a white background using Microsoft Visual Basic 6.0. The experiment took place in a quiet, well-lit room. Participants were tested in individual sessions that lasted about 20 min. Subjects were seated next to the experimenter so that both the experimenter and the subject had roughly the same point-of-view. Participants were given six practice trials: the first two practice trials illustrated the concept of "exactly the same as" while the rest familiarized the children with the experimental procedure. For example, in one of the first two practice trials, subjects were shown a picture of a tiger at the top-center of the screen (target), with pictures of the target tiger, a cow, and a snake on the bottom part of the screen. Participants chose from the test items by pointing to one of the test items, and were given feedback about their choice. In the remaining 4 practice trials, geometric relations were used to familiarize children with the actual experiment. For example, in one of these practice trials participants were shown an oval split in two by color (blue on the one side and white on the other) and were told "look at this carefully and help me find one exactly like it" in the No Label condition while they were told "the blue is on the right of/prettier than the white" in the language conditions. After the target disappears, followed by a 1 s delay, children chose among the target replica, its reflection and a distracter for the "match" by pointing to one of the test items, and were given feedback about their choice. They were encouraged to look at all three test items before responding.

The six different target stimuli were each presented four times for a total of 24 trials, randomly ordered. Test item location was completely randomized such that across trials the correct item was at each of the three locations an equal number of times. Participants' accuracy was the dependent variable.

2.3. Post-tests

At the end of the delayed-matching task, children were given post-tests to evaluate their knowledge of the spatial and non-spatial labels. In the No Label and Directional conditions, participants were tested on knowledge of "left" "right" "top" and "bottom". In the Production test, a 3 cm × 3 cm square was displayed in the middle of a computer screen with a smiley face placed next to one of its sides (left, right, top, bottom; Fig. 3). A smiley face was placed at each side eight times, for a total of 32 trials. On each trial, the subject was asked 'Where is the smiley face to the square?'. They had to answer by producing a term or phrase expressing the location of the smiley (e.g., left, right etc.). In the Comprehension task children were given a 20 cm × 14 cm paper with a 3 cm × 3 cm square drawn at the center. They were asked to point to 'the left/right/top/bottom' side of the square. Each term was tested eight times, for a total of 32 trials. For each subject, the Production task was conducted before the Comprehension task.

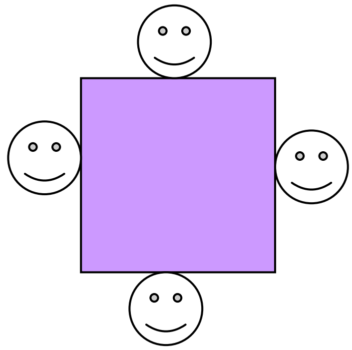

Fig. 3. Left/right/top/bottom production post-test. To assess children's knowledge of the terms "left", "right", "top" and "bottom", children were shown a square with a smiley face in one of the four locations. Participants were shown the square-smiley face combination and were asked "Where is the smiley face?" and were prompted with "The smiley face is. . ..". See text for parallel comprehension test.

Children who participated in the Prettier condition were given a comprehension task to test their knowledge of the non-reversibility of the non-spatial asymmetric predicates. A puppet named Bobo was introduced to the child. The child was told that Bobo knows a lot about Disney characters and that Bobo was going to tell him/her some things that s/he has to "listen very carefully and remember". Then, the children were shown two Disney characters on a computer screen. Bobo (a hand held puppet with the voice of the experimenter) then made a statement about the two characters. For example, "This is Goofy and this is Mickey Mouse. You know what? Both Goofy and Mickey Mouse are really nice. But guess what, I think Mickey Mouse is *nicer* than Goofy. Remember, I think Mickey Mouse is *nicer* than Goofy." Then, the experimenter asked the child "Does Bobo think that Goofy is nicer than Mickey?" and then "Does Bobo think that Mickey is nicer than Goofy?" A 'yes' or 'no' answer was encouraged after each question. The four terms used in the Prettier label condition (nicer, prettier, brighter, lighter) were each mentioned by Bobo in 4 different scenarios for a total of 16 trials. The two questions in each trial were structured to test whether the children understood the non-reversibility of the predicates, e.g. the relationship defined in a statement like "X is nicer than Y". If so, the child should say "yes" when asked if "X is nicer than Y" but "no" when asked if "Y is nicer than X". Each trial was given a score of 1 if the child said "yes" to one of the questions and "no" to the other (i.e., the child understood the nonreversible nature of the relationship defined). The trial was scored 0 if a child said "yes" or "no" to both of the questions in a trial.

2.4. Results and discussion

Percents of each choice across the three conditions in the matching task are shown in Fig. 4 (Experiment 1A, left panel). Consistent with our previous work, 4 year-old children chose the Correct test item more often than chance

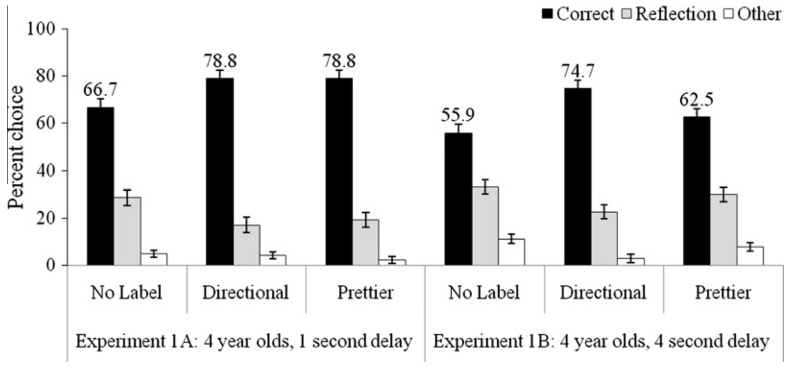

Fig. 4. Experiment 1 results. Mean percent choice of Correct, Reflection and Other for 4 year-olds in Experiment 1A (1 s delay) and 1B (4 s delay). Error bars indicate a standard error of the means.

(33%) in each of the label conditions: No Label, $t(11) = 8.27$, $p < .001$, M percent Correct = 66.7, Directional, $t(11) = 15.00$, $p < .001$, M percent Correct = 78.8, and Prettier, $t(11) = 13.2$, $p < .001$, M percent Correct = 78.9. Children also chose significantly more Reflections than Other distracters in each condition, No Label, Wilcoxon $t(11) = 2.29$, $p < .01$, Directional, Wilcoxon $t(11) = 2.99$, $p < .01$, and Prettier, Wilcoxon $t(11) = 3.12$, $ps < .01$.

Children in the No Label condition accurately represented the internal geometry of the square (Vertical, Horizontal, or Diagonal split) with combined choices of the Correct (67%) and Reflections (28%) accounting for 95% of the responses. Errors largely involved incorrect assignment of color to location, with 85% of the errors Reflections. Thus children in this condition had difficulty maintaining the conjunction of color and location in the task, replicating previous work with children (Dessalegn & Landau, 2008; Hoffman, Landau, & Pagani, 2003) and abundant work with adults.

Did the labels help improve children's memory for color-location conjunction? A mixed analysis of variance with Condition (No Label, Directional, Prettier) and Target type (Vertical, Horizontal, Diagonal) showed a significant effect of Condition, $F(2,33) = 3.87$, $p < .05$ but no effect of Target type $F(2,66) = 1.63$, $p > .05$ (Mean percent correct for the Diagonal = 72.91; Horizontal = 79.17; Vertical = 72.22), nor interaction between Condition and Target type. Planned-comparisons showed that children performed significantly better in the Directional and Prettier conditions compared to the No Label condition, both $ts(22) > 3.0$, $ps < .05$. There was no difference between the Directional and Prettier conditions, $t(22) < 1$, $p > .05$. Thus, the linguistic instructions containing asymmetric predicates – including *prettier, nicer, brighter, lighter, left, top, bottom and right* – helped children remember the target better. This improvement was observed regardless of whether the labels were spatial or non-spatial.

2.5. Post-test results

Was performance in the delayed-matching task related to children's long-term knowledge of the terms used in the instructions? If so, then, in the No Label and Directional conditions, children who knew the terms left/right/top/bottom should have performed better than those who did not know the terms. In the Prettier condition, children who knew that the terms prettier/nicer/brighter/lighter were asymmetric and non-reversible should have performed better than those who did not.

Looking at the Production data for directional terms, children in the No Label condition correctly produced both "top" and "bottom" on 100% of the trials, and in the Comprehension test they placed the X in the correct location on 100% of the trials. Thus children knew the spatial meanings of "top" and "bottom". However, the same children produced "left" and "right" correctly only on 58% and 48% of the trials, respectively. Similarly, they placed the X in the correct "left" and "right" locations on only 40% and 48% of the trials. That is, although these children knew the meanings of "top" and "bottom" their knowledge of "left" and "right" was quite limited.

Children in the Directional condition showed a similar pattern. They correctly produced both "top" and "bottom" on 100% of the trials and placed the X in the correct "top" and "bottom" location on 99% and 100% of the trials respectively. On the other hand, the children correctly produced both "left" and "right" on only 66% of the trials, and they correctly placed the X on the "left" and "right" side on only 60% and 62% of the trials. Like the 4 year-olds in our previous work, in both the No Label and Directional conditions, the production of left/right was reliably worse than top/bottom, $ts(10) > 3.40$, $ps < .01$, as was Comprehension, $ts(10) > 3.60$, $ps < .01$.[4] Thus 4 year-olds had a solid long-term knowledge of the terms "top" and "bottom" but quite fragile knowledge of "left" and "right". Their errors on this task suggest that children of this age do know that the terms left/right apply to the horizontal axis, but do not know which end of the horizontal axis is "left" and which is "right". Specifically, almost all of the errors observed in the production and comprehension tasks were confusions of "left" with "right" but not, e.g., "left" with "top". In both the No Label condition and Directional Label conditions 100% of errors in

[4] One participant in No Label, one participant in the Directional, and 2 children in Prettier condition did not complete these post-tests due to time constraint or exhaustion.

Production and Comprehension were axis-appropriate errors (e.g., saying or indicating "left" instead of "right").

To evaluate whether children's long-term knowledge of the spatial terms was related to performance on the delayed-matching task, we carried out a correlation between children's accuracy in the matching task and their accuracy for left/right in the Production and Comprehension tasks.[5] Regardless of the condition, children's production of the left/right was not significantly correlated with their matching accuracy in the Directional condition (Pearson's $r(10) = .53$, $p > .05$ or in the No Label condition, $r(10) = .10$, $p > .05$.[6] The same was true for children's comprehension of the spatial terms with one exception – children's accuracy in the Directional condition matching task was significantly correlated with their comprehension of left/right, Pearson's $r(10) = .62$, $p = .04$. Looking at the scatter plot for this dataset it was clear that this effect was a result of two participants who had high scores on the comprehension task and matching task (indeed, removing even one of these participants removes the significant effect, Pearson's $r(9) = .54$, $p = .11$). Thus the findings suggest that children were not entirely relying on their long-term knowledge of the spatial terms in this task, a finding that replicates our previous finding (Dessalegn & Landau, 2008). That is, how well they know left/right is not directly related to their memory for the color-location conjunctions. We will return to this point in Section 8.

On the other hand, 4 year-olds were at ceiling in the Prettier post-test with a percent correct of 90%, showing that these children knew the oppositional nature (i.e., non-reversibility) of the relationship defined by the terms. Because of the lack of variation in the Prettier post-test data a correlation was not carried out between performance on the delayed-matching task and their knowledge of the non-spatial asymmetric terms.

The findings thus far show that 4 year-olds showed performance improvement when the target was accompanied by sentences containing spatial ("the yellow is *on the left* of the black") or non-spatial ("the yellow is *prettier than* the black") asymmetric predicates. This supports the idea that what is crucial for the observed language effect – enhanced representation and/or memory for the color/location conjunctions – is the abstract semantic properties of the predicates.

These findings support the surprising idea that the two types of terms – left/right and prettier – might have exactly equivalent effects on enhancing children's memory for the visual stimuli. Although the findings for *prettier* certainly suggest that the effects are not restricted to spatial predicates like left/ right, they do not necessarily suggest an exact equivalence. One possibility is that these two types of predicates have equivalent effects for the timeline we offered children – that is, with a 1-s delay between instruction and test – but that over longer timelines, the effects might be quite different. For example, the effects of instruction with left/right might be more resistant to decay than *prettier* simply because they express a spatial relationship, which is what is needed, in the end, to successfully match the target to test items. In Experiment 1B, we address this possibility by increasing the delay between target and test items from 1 s to 4 s. If children are truly treating the spatial (*right/left*) and non-spatial (*prettier, nicer*) asymmetric and non-reversible predicates as members of the same equivalence class (disregarding the additional information carried by the spatial predicates), their performance in the two conditions should be affected similarly when the delay is increased. Alternatively, if the spatial labels are helping children create a more robust representation in the Directional condition compared to the Prettier condition, then performance in the two conditions should be affected differently when the task is made harder by increasing the delay.

3. Experiment 1B

3.1. Participants

A new group of 36 4 year-olds ($M = 4;4$, Range = 4;0–4;11) participants were randomly assigned to either a No label ($n = 12$), Directional label ($n = 12$), or Prettier label ($n = 12$) condition.

3.2. Design, stimuli and procedure

These were exactly the same as in Experiment 1A with one exception – the delay between target offset and test-item presentation was increased to 4 s.

3.3. Results and discussion

Children again chose the Correct item more often than chance (33%) in each of the label conditions, all $ts > 5$, all $ps < .001$ (M percent correct for No Label = 55.9, Directional = 74.7, and Prettier = 62.5). Moreover children also chose more Reflections than Other distracters in each condition, all $ts > 3$, all $ps < .01$ (Fig. 4, right panel).

A mixed analysis of variance with Condition (No Label, Directional, Prettier) and Target type (Vertical, Horizontal, Diagonal) as variables showed a significant effect of Condition, $F(2,33) = 7.12$, $p < .01$ but no effect of Target type or interaction between Condition and Target type. In contrast to the findings of Experiment 1A, planned-comparisons showed that children performed significantly better in the Directional condition compared to the Prettier and No Label conditions, both $ts(22) > 2.6$, $ps < .02$. There was no difference between the No Label and Prettier conditions, $t(22) = 1.1$, $p > .05$.

Comparing performance in Experiment 1A with performance in Experiment 1B, overall performance in Experiment 1A (1 s delay, M percent correct = 74.8) was significantly higher than in Experiment 1B (4 s delay, M percent correct = 64.4), $F(1,70) = 10.19$, $p < .001$. Planned-comparisons showed that there was a significant performance decline in the Prettier condition in Experiment 1B compared to Experiment 1A, $t(22) = 3.10$, $p < .01$, and a marginal drop in the No Label condition, $t(22) = 1.86$,

[5] Children were at ceiling for *top/bottom*, thus only left/right performance was entered into correlations.

[6] We suspect that this relatively high but non-significant, correlation is being driven by a few point and the small subject sample (please see the supplemental materials for the correlation plots).

$p = .076$. There was no difference between Experiment 1A and 1B in the Directional condition, $t(22) < 1$, $p > .05$.

Thus, although sentences containing spatial and non-spatial asymmetric predicates seemed to facilitate performance equally well when children had to maintain the visual representation for only 1 s, instructing with these different predicate classes leads to different results when the delay is increased to 4 s. This finding suggests that 4 year-old children represent the semantic difference between the sentence containing spatial and non-spatial asymmetric predicates, and are better able to maintain an advantage when the semantics encodes the spatial relationship.

In sum, Experiments 1A and 1B show that 4 year-olds can use abstract semantic properties of predicates to enhance their memory for symmetric visual stimuli, but that effects of the abstract properties of non-reversible predicates as a group – that is, without spatial content – are relatively short-lived.

Given both the enhancing effects of spatial and non-spatial asymmetric predicates and their limitations among 4 year-olds, we now ask whether these effects are also present earlier in development. We do this by asking whether the observed advantages of asymmetric and non-directional labels are present among 3 year-olds.

4. Experiment 2A

4.1. Participants

Thirty-six 3 year-olds ($M = 3;3$, Range = 3;0–3;8) were randomly assigned to either a No label ($n = 12$), Directional label ($n = 12$), or Prettier label ($n = 12$) condition.

4.2. Design, stimuli and procedure

These were exactly the same as Experiment 1A.[7]

4.3. Results and discussion

Three year-olds chose the Correct item more often than chance (33%) in each of the conditions (*M* percent correct in No Label = 54.5, Directional = 60.8, Prettier = 52.4, all $ts > 4$, all $ps < .01$), showing that they understood the task. Moreover they chose significantly more Reflections than Other distracters in each condition, all Wilcoxon $ts > 2.5$, all $ps < .01$, with combined choices of the Correct and Reflections accounting for over 85% of choices. This shows that, just like 4 year-olds, 3 year-olds have a robust representation of the internal geometry of the target but err on the color-relative location conjunction, resulting in an incorrect assignment of color to location, with 78%, 70%, and 72% of their total errors being reflection errors in the No Label, Directional, and Prettier conditions respectively (Fig. 5, left panel).

A 3-Condition (No Label, Directional, Prettier) × 3-Target type (Vertical, Horizontal, Diagonal) mixed ANOVA showed no significant effect of Condition, $F(2,33) = 1.07$, $p > .05$, Target type $F(2,66) = 1.94$, $p > .05$ (Mean percent Correct for Diagonal = 55.21; Horizontal = 61.11; Vertical = 51.39), or interaction, $F(4,66) < 1$, $p > .05$. Thus, although the visual representation of the target is fragile in the case of 3 year-olds – just as with 4 year-olds – 3 year-olds fail to show any improvement in their memory for the target using either the directional terms (top/bottom/left/right) or the non-spatial asymmetric terms (prettier/nicer/lighter/brighter). That is, the linguistic instructions had no effect on the 3 year-olds' performance. Finally 3-year-olds' performance on the delayed-matching task was not related to their performance on the post-tests.[8]

Experiment 2A showed that although 3 year-olds understood the task and performed above chance in all conditions, they did not use the *left/right* or *prettier* instructions to improve their performance. One possibility is that the duration of the delay was simply too long for 3 year-olds to hold onto this linguistic information. If so, then any effects of language could not be observed with the time-line we used. To address this possibility, we ask in Experiment 2B whether any effects of language are observed with a shorter delay between the disappearance of the target and appearance of the test items.

5. Experiment 2B

5.1. Participants

A new group of 36 3 year-olds ($M = 3;3$ Range = 3;0–3;6) was recruited and randomly assigned to a No Label ($n = 12$), Directional label ($n = 12$), or Prettier label ($n = 12$) condition.

5.2. Design, stimuli and procedure

These were exactly the same as Experiment 2A, except that the delay between target offset and appearance of the test items was reduced from 1 s to .5 s.

[7] The post-tests were administered to all of the participants in all of the experiments reported here. Because the results are very similar and to conserve space, only one set of results is reported. For the full data please see Dessalegn, 2009.

[8] Across the No Label and Directional conditions 3-year-old children correctly produced "top" and "bottom" on more than 95% of the trials, and in the Comprehension test they placed the X in the correct location on more than 96% of the trials. On the other hand, children in the No Label condition correctly produced both "left" and "right" on only 50% of the trials and in the Comprehension task, they placed the X in the correct "left" and "right" location on 57% and 62% of the trials. In the Directional condition children produced "left" and "right" on 35% and 42% of the trials and in the Comprehension task they placed the X in the correct "left" and "right" location on 49% and 34% of the trials. As with the 4 year-olds, although 3-year-olds knew the meanings of "top" and "bottom", their knowledge of "left" and "right" was quite limited. Specifically, the production of left/right was reliably worse than top/bottom, $ts > 4.0$, $ps < .01$, as was their comprehension, $ts > 3.98$, $ps < .01$. Like 4 year-olds, 3-year-olds knew that the terms left/right apply to the horizontal axis but they did not know which end of the horizontal axis was "left" and which was "right" – more than 89% of their errors showed correct use of the horizontal axis without correct use of the direction within the axis. A correlation between the 3 year-olds' accuracy in the delayed-matching task with their accuracy in the Production and Comprehension tasks showed no significant correlations regardless of condition, No Label condition, all $rs < .60$, $ps > .05$ and Directional condition all $rs < -.03$, $ps > .05$. Nine of the 12 participants in the Prettier condition were given the Prettier post-test. This was a difficult task for 3-year olds, with an overall percent correct of 34. There was no significant correlation between performance on the delayed-matching task and performance on the prettier post-test, $r(8) = -.69$, $p > .05$.

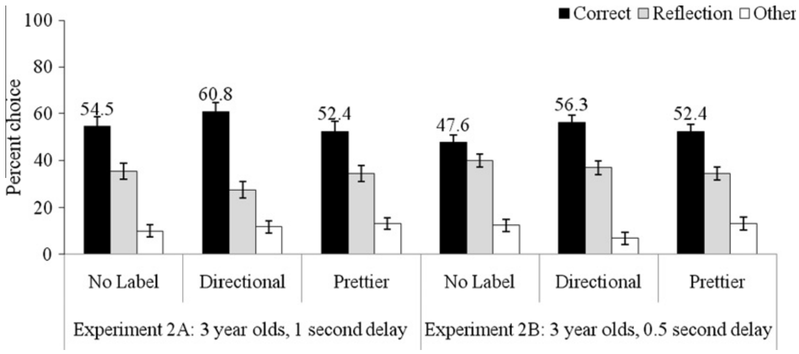

Fig. 5. Experiment 2 results. Mean percent choice of Correct, Reflection and Other for 3 year-olds in Experiment 2A (1 s delay) and 2B (0.5 s delay). Error bars indicate a standard error of the means.

5.3. Results and discussion

Results replicated the pattern of performance for 3 year-old children observed in Experiment 2A. That is, there were still no effects of linguistic instruction beyond the simple instruction to "Look at this!". Percent correct was significantly better than chance in all conditions, all $ts > 4$, all $ps < .01$, and there were significantly more Reflection errors compared to Other errors, all Wilcoxon $ts > 2.8$, all $ps < .01$. Importantly, there were no reliable effects of instruction (*M* percent correct No Label = 47.6, Directional = 56.3, and Prettier = 52.4). A mixed ANOVA with Condition (No Label, Directional, Prettier) and Target type (Vertical, Horizontal, Diagonal) showed no significant effect of Condition, Target type or interaction (Fig. 5, right panel). Directly comparing 3 year-olds' performance in Experiment 2A and 2B, there was no difference between overall accuracy in the two experiments *M* percents correct = 55.9, 52.1, respectively, $F(1,70) = 1.58$, $p > .05$.

Error patterns in both experiments were virtually identical. These findings strongly suggest that the linguistic instructions we provided did not affect children's representation and memory for the target items. This points to the possibility that, at age 3, visual representations of the target and the linguistic representations of the verbal instruction operate rather independently of each other, and that the ability to combine useful and useable information from the two modalities develops over time.[9] If so, then we would want to know whether the effects observed among 4 year-olds are the end of the developmental line, or whether the interactions between vision and language in our task change further. In order to find out, we carried out Experiment 3, in which we examined the dynamics of the interactions among 6 year-olds using the same instructional contexts, first with a 1 s delay (Experiment 3A), and then with a 4 s delay (Experiment 3B).

[9] We will come back to this point in Section 8 and suggest that this may only be true for very specific types of interactions since there is evidence for other language–vision interactions that are well in place by this age.

6. Experiment 3A

6.1. Participants

Thirty-six 6 year-olds ($M = 6;3$ Range = 6;0–6;8) were randomly assigned to a No Label ($n = 12$), Directional label ($n = 12$), or Prettier label ($n = 12$) condition.

6.2. Design, stimuli and procedure

These were exactly the same as Experiment 1A and 2A, with a 1 s delay.

6.3. Results and discussion

Six-year-olds were significantly above chance in choosing the Correct item in each condition, *M* percent correct in No Label = 93.1, Directional = 96.2, Prettier = 83.7; all $ts > 13$, all $ps < .01$. They also represented the internal geometry of the square accurately with combined choices of the Correct and Reflections accounting for more than 98% of the responses across the conditions (see Fig. 6, left panel).

A mixed analysis of variance showed a significant effect of Condition, $F(2,33) = 6.59$, $p < .01$ but no effect of Target type $F(2,66) < 1$, $p > .05$ (Mean percent Correct for Diagonal = 90.28; Horizontal = 92.71; Vertical = 89.93), nor interaction, $F(4,66) < 1$, $p > .05$. The main effect of condition was a result of children's better performance in the No Label and Directional label conditions compared to the Prettier condition. Planned-comparisons showed no significant difference between the No Label and Directional conditions, while performance in both the No Label condition and the Directional condition were significantly better than the Prettier condition ($t(11) = 3.06$, $p < .05$, $t(11) = 3.54$, $p < .001$, respectively).[10]

[10] Six year-olds were at ceiling in the production and comprehension post-tests for the directional terms as well as in the Prettier post-test. On the production task, children were 91% correct for the terms left and right, and they were 100% correct for top and bottom. Similarly, in the Comprehension task, children were 93.7% correct for left/right and 98.6% correct for top/bottom. In the Prettier task, children were 100% correct. Because there was not enough variation in their production and comprehension data, correlations between production and comprehension results and matching task performance were not carried out.

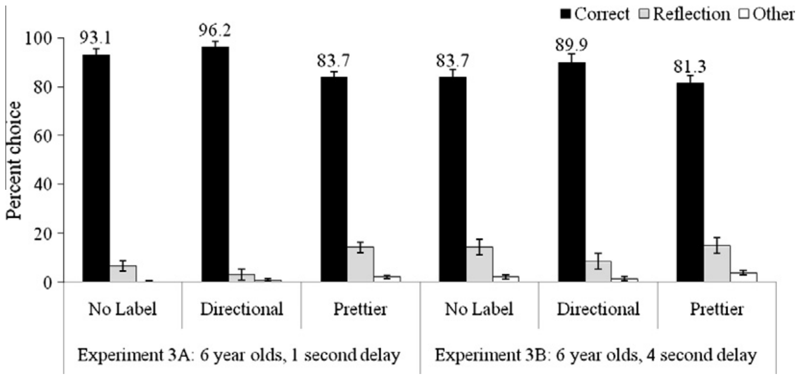

Fig. 6. Experiment 3 results. Mean percent choice of Correct, Reflection and Other for 6 year-olds in Experiment 3A (1 s delay) and 3B (4 s delay). Error bars indicate a standard error of the means.

Thus for 6 year-olds there was no difference between the baseline (No Label) condition and the Directional label conditions. These findings might be taken to suggest that 6 year-olds, like adults in previous studies (e.g., Dessalegn, 2009; Stefurak & Boynton, 1986), automatically deploy language—in this case, directional terms – to augment their performance in the delayed-matching task. Under such a story, the performance decline observed in the Prettier condition is a result of the non-spatial labels (e.g., "prettier") interfering with the automatic labeling of targets using spatial labels by children. That is, providing the labels in the Prettier condition might have produced effects similar to a language interference task.

Alternatively, it is possible that in the baseline condition (No Label condition) children were able to form a robust representation of the conjunctions (enough to last 1 s of delay) without automatically recruiting language and that while the Directional labels produced an added benefit, this was not detectable due to ceiling performance in both cases. In Experiment 3B we attempt to distinguish between these possibilities by extending the delay to 4 s. If language is being automatically used in the No Label condition, then increasing the delay in the task should affect the No Label and Directional conditions similarly. But, if children are forming a more robust representation in the Directional condition compared to the baseline condition, then children in the Directional label condition should show better performance than those in the No Label condition at the longer 4 s delay.

7. Experiment 3B

7.1. Participants

A new set of 36 6 year-olds ($M = 6;2$ Range = 6;0–6;6) participated in this experiment. Children were randomly assigned to a No Label ($n = 12$), Directional label ($n = 12$), or Prettier label ($n = 12$) condition.

7.2. Design and stimuli

These were exactly the same as Experiment 3A, except that the delay between target offset and test-item presentation was increased to 4 s.

7.3. Results and discussion

Six year-olds chose the Correct item more often than chance (33%) in each of the label conditions $ts(11) > 12$, $ps < .001$ (Fig. 6, right panel). A mixed analysis of variance with Condition (No Label, Directional, Prettier) and Target type (Vertical, Horizontal, Diagonal) showed no significant effect of Condition, Target type, nor interaction. Unlike the results of Experiment 3A, the No Label and Directional Label condition were not different from Prettier Condition. Moreover, comparison of Experiment 3A and 3B showed a significant overall performance decline in Experiment 3B (M percent correct = 85.0) compared to Experiment 3A (M percent correct = 91.0), $F(1,70) = 5.18$, $p < .04$; planned contrasts showed a significant decline in performance in the Directional, $t(22) = 2.16$, $p < .05$, and No Label, $t(22) = 2.19$, $p < .05$, conditions. Thus, the No Label and Directional conditions seem to be subject to similar restrictions, including decay rate. These finding strongly suggest that 6 year-olds in the No Label condition automatically deploy language to form a robust representation which they can maintain in visual working memory more easily as has been suggested for adults.

8. General discussion

We framed our paper using several questions about the interaction between vision and language. Does language change the content and structure of the visual representation with which it interacts? Is the effect of language time bound, seen only in the confines of a specific task or does it have a more long-term or permanent effect? What properties of language are relevant for the observed effect of language? Finally are there developmental changes in whether (and the degree to which) language interacts with nonlinguistic cognition? Our previous work suggested that language does not change the target visual representation, because errors of confusion for color /location combinations occur systematically and regularly among children and adults without the intervention of specific linguistic instructions. That work also suggested that the effects are momentary – seen within the confines of the task. Our present studies reinforce this conclusion, at least among 4 year-olds, but add to our understanding of the timeline within which the effects occur. Moreover, our findings suggest that the effects depend on highly abstract representations of the linguistic instructions, and that these effects develop between ages 3 and 6. We discuss each of these points below.

8.1. It's momentary

Our findings reinforce our earlier conclusion about momentary effects, but also suggest some new facts about the timeline under which the effects operate. First, we replicated our earlier effect showing that, with a 1-s delay, 4 year-olds were better able to match the target item if they had been instructed with sentences containing *left/right*, but that moments after the task was complete, the same 4 year-olds showed incomplete knowledge of the directionality of these terms. As in our previous results, there was no reliable correlation between children's knowledge of directionality of the terms and their performance on the matching task. But we also found that after a 4-s delay, the profile of performance changed over the different linguistic conditions. There was a general decrement in performance with the longer delay, but perhaps more interesting, the directional/ spatial predicates (*left/right*) retained their influence while the asymmetric non-spatial predicates (*prettier*) now lost their power to enhance memory, dropping to a level similar to the No Label instruction. This suggests that 4 year-olds represent the difference between the two types of predicates with the spatial predicates (most relevant for the task) helping 4 year-olds to create and retain a more robust representation even over a 4-s delay.

The timeline for 6 year-olds appears to be quite different, with ceiling performance for both No Label and Directional Label at 1 s, and a drop in both conditions at 4 s. Unlike these two conditions, the *prettier* condition elicited poor performance among 6 year-olds in both the 1 and 4-s conditions. At the age of 6, of course, children were at ceiling in the production and comprehension tasks measuring their knowledge of *left/right*, so we do not see the same dramatic difference between performance on the matching task and the post-tests as we observed with 4 year-olds. We suspect that by age 6, children may be automatically encoding the stimuli using spatial language, which can then be used to span the 4-s delay. This would explain why the *prettier* condition never had the same enhancement effects for 6 year-olds as it did for 4 year-olds.

8.2. It's abstract

Among the most striking of our findings is the finding that 4 year-olds can benefit from sentences containing spatial asymmetric and non-reversible predicates (*left/right*) as well as non-spatial asymmetric and non-reversible predicates (e.g., *prettier*). At the same time, the non-spatial predicates did not enhance performance for 6 year-olds. These findings suggest that at age 4, an abstract representation of asymmetry and nonreversibility as part of a predicate's structure, is used to enhance children's memory of visual stimuli, but that by age 6, these properties alone – without spatial content – are no longer useful. At this point, it seems likely that only predicates expressing *spatial* asymmetry/non-reversibility (e.g., left) may be useful to children in this task; even this conclusion, though, should be made with caution, since the 6 year-olds also performed well without *left/right*. Taken together with our previous findings (Dessalegn & Landau, 2008), our current findings suggest that in cases where vision fails to form a robust representation, linguistic information – specifically, abstract representations of a predicate's structure – can help form and/or maintain a more robust representation. Furthermore, the findings suggest that, at age 4, the effect of language depends on highly abstract semantic properties of language. At the age of 4, children attend to and use an abstract semantic notion of "non-reversibility" to help them form a robust representation of visual feature conjunctions, while by age 6 a more specific, combination of semantic features (i.e. spatial + nonreversible) is required to improve performance.

8.3. It develops

There was a clear developmental pattern in the effects of linguistic instruction on enhancing children's ability to represent and remember the color/ location combination of the visual stimuli they saw. Three year-olds showed no effects at all, 4 year-olds showed effects modulated with linguistic instruction and delay, and 6 year-olds showed only effects of interference with the *prettier* instructions. As we discussed, we suspect that 6 year-olds in the No Label condition may have been automatically encoding the stimuli with asymmetric spatial predicates such as *left/right*, and that this accounted for the lack of an advantage when they actually heard instructions with these predicates. By contrast, when they heard *prettier*, they may have wondered why, pragmatically, the experimenter was commenting on the aesthetic value of the target, and this may have caused some distraction, leading to poorer performance. Importantly, all age groups showed that they can stably represent the internal geometry of the target blocks; they almost never confused a left/right split with a top/bottom or diagonal split. Both 3 and 4 year-olds also showed a fragile memory for the conjunction of color and location, making reflection errors across all conditions and timelines.

The striking difference over age suggests the following developmental scenario: At 3 years of age, representation of the visual stimulus may not engage language at all, while at 6 years of age, children are likely to automatically recruit language to help them better remember the color/ location conjunctions. This pattern, taken together with previous developmental work (Duncan, Whitney, & Kunen, 1982; Hagen & Kingsley, 1968; Reese, 1975) suggests that for a certain class of vision-language interactions, vision and language become increasingly linked across development such that, in adulthood language is automatically activated during visual–perceptual and visual memory tasks (see, e.g., Bruner, Busiek, & Minturn, 1952; Carmichael, Hogan, & Walter, 1932; Miller, 1994; Stefurak & Boynton, 1986).

Similar suggestions about the developmentally emerging gradual linking of vision and language have been made by others (e.g., Flavell, 1970; Flavell et al., 1966;

see also Baddeley et al., 1998; Hagen & Kingsley, 1968; Loewenstein & Gentner, 2005; Reese, 1975). For example, Flavell et al. (1966) tested children's immediate and delayed memory for object identity and serial order. In the task, an experimenter touched several objects (of 7 objects total) in a particular order, which they called serial order. The 7 objects were then spatially rearranged and children were asked to touch the same objects in the same order immediately after the spatial rearrangement of the object (or 15 s after the rearrangement). Children younger than about 5 years (kindergartners) did not automatically produce linguistic labels in this task, while 2nd and 7th graders did. Similarly, Reese (1975) found that children between the ages of 2;8 and 4;7 failed to produce useful verbal descriptions of visually presented spatial relationships in a memory task, while 5–6 year-olds automatically did so. Hagen and Kingsley (1968) conclude that "spontaneous use of rehearsal strategies is seldom employed by children under 6 years" (pg. 6), and Duncan et al. (1982) concluded that "...results illustrate the increasing interdependence of the verbal and visual systems with age" (pg. 1215). Our results are consistent with these conclusions, and suggest that vision and language become gradually more interdependent across development.

The conclusion of developmentally increasing interactions between vision and language may seem to be at odds with much work in developmental psychology that has shown such interactions early in development (e.g., Fisher, Hall, Rakowitz, & Gleitman, 1994; Fisher & Snedeker, 2002; Landau & Stecker, 1990; Soja, Carey, & Spelke, 1991; Subrahmanyam, Landau, & Gelman, 1999). For example, Fisher and Snedeker (2002) have shown that 21 and 26 months old children chose a two-actor scene if they heard sentences containing two nouns (e.g., "she is blicking him") and chose a one-actor scene when they heard a sentence containing only one noun (e.g., "she is blicking"), showing sensitivity to the mapping between the number of perceived entities and the number of heard nouns representing arguments.

We think such interactions may be quite different from those that are the focus of our work. In our studies, we have examined the role of language in enhancing memory and representation in cases where the visual system is inherently fragile. By contrast, the interactions discussed in Fisher and Snedeker (2002) and in many similar works illustrate interactions between language and perception where the perceptual representations are quite robust and are present quite early in development, even pre-linguistically. Our findings suggest a role for language that is more powerful than these cases – a role that allows the perceiver to move beyond visual representations and store information that would otherwise be fragile or degraded.

There remain several open questions about the nature of the mechanism we have described. One is the mystery of why 4 year-olds benefited from the directional predicates, even though they did not have complete knowledge of the directionality inherent in the predicates. We suggest that children's knowledge of the spatial terms was "updated" to add directional information in the immediate context of the task. That is, when the child hears "the red is on the left" while viewing the vertically split square, he might simply note that the term *left* refers to the red part, thereby temporarily storing the correct directional entry for *left*. This correct representation of *left* can then be used to bridge the 1-s delay, leading to a correct match. Later, when children are given the production and comprehension post-tests (a mere 10–15 min later) children have lost that representation and continue to err on the directionality of the terms.[11]

A second open question concerns the nature of the representation resulting from seeing the visual stimulus and hearing the linguistic instruction. One possibility is that language modifies the existing fragile visual representation – that is, the visual representation is updated to create a new representation – with an asymmetry where none existed before. A second possibility is that language serves as an "annotator" to the visual representation, resulting in a hybrid representation containing elements of both the original visual representation and the linguistic instruction. We suspect that these two possibilities maybe difficult to distinguish experimentally.

To conclude, our work has provided evidence for a developmentally emergent interaction between vision and language that is momentary, abstract, and undergoes significant development between ages 3 and 6. Laying out the precise mechanisms whereby language interacts with vision advances our understanding of the role of language in the human mind, and moves us beyond simplistic questions about whether or not language affects thought. Understanding the nature of language–thought interactions requires not just a listing of observed language-based shifts in behavior, but also an articulation of the conditions under which these shifts occur and mechanistic explanations of how they occur. We believe that focusing on specific mechanisms of interaction between language and cognition – rather than simply enumerating effects and non-effects of language on cognition – can offer us deeper understanding of the powerful role that human language plays in human cognition.

Acknowledgments

This research was supported by an NRSA Grant F31HD056629 from the National Institute of Child Health And Human Development to BD, a Provost's Career Enhancement Postdoctoral Scholarship from the University of Chicago to BD, and NINDS RO1 NS 050876 to BL.

[11] An alternative possibility is that statements like "The yellow is on the left of/prettier than the black" help because the sentence picks out and labels one part of the target. Thus even in cases where children do not know the full meaning of the predicate the sentence helps because it labels a part. This cannot be the whole story because, as reported in Dessalegn and Landau (2008), children hearing other sentences that have this property of picking out a part, e.g., "the yellow is touching the black", fail to show the same performance improvement.

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.2013.02.003.

References

Baddeley, A., Gathercole, S., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review, 105*(1), 158–173.
Barner, D., Inagaki, S., & Li, P. (2009). Language, thought, and real nouns. *Cognition, 111*, 329–344.
Bruner, J., Busiek, R., & Minturn, A. (1952). Assimilation in the immediate reproduction of visually perceived figures. *Journal of Experimental Psychology, 44*(3), 151–155.
Carey, S. (2009). *The origin of concepts*. New York: Oxford University Press.
Carmichael, L., Hogan, H. P., & Walter, A. A. (1932). An experimental study of the effect of language on the reproduction of visually perceived form. *Journal of Experimental Psychology, 15*(1), 73–86.
Davidoff, J. (2004). Coloured thinking. *The Psychologist, 17*(10), 570–572.
Davidoff, J., Davies, I., & Roberson, D. (1999). Colour categories in a Stone-Age tribe. *Nature, 398*, 203–204.
Dessalegn, B. (2009). *A vision of language: The mechanism, development and consequence of an interaction between language and vision*. Retrieved from ProQuest® Dissertations & Theses (publication number: 3357163).
Dessalegn, B., & Landau, B. (2008). More than meets the eye: The role of language in binding and maintaining feature conjunctions. *Psychological Science, 19*(2), 189–195.
Duncan, E., Whitney, P., & Kunen, S. (1982). Integration of visual and verbal information in children's memories. *Child Development, 53*(5), 1215–1223.
Egeth, H., & Smith, E. (1967). Perceptual selectivity in a visual recognition task. *Journal of Experimental Psychology, 74*(4), 543–549.
Fisher, C., Hall, D. G., Rakowitz, S., & Gleitman, L. R. (1994). When it is better to receive than to give: Syntactic and conceptual constraints on vocabulary growth. *Lingua, 92*, 333–375.
Fisher, C., & Snedeker, J. (2002). Counting the nouns: Simple sentence-structure cues guide verb learning in 21-month-olds. In *Paper presented at the Twenty-seventh Boston University Conference on Language Development*, Boston, MA, November 2002.
Flavell, J. H. (1970). Developmental studies of mediated memory. *Advances in Child Development and Behavior, 5*, 181–211.
Flavell, J. H., Beach, D. R., & Chinsky, J. M. (1966). Spontaneous verbal rehearsal in a memory task as a function of age. *Child Development, 37*(2), 283–299.
Frank, M. C., Everett, D. L., Fedorenko, E., & Gibson, E. (2008). Number as a cognitive technology: Evidence from Pirahã language and cognition. *Cognition, 108*, 819–824.
Frank, M. C., Fedorenko, E., Lai, P., Saxe, R., & Gibson, E. (2012). Verbal interference suppresses exact numerical representation. *Cognitive Psychology, 64*, 74–92.
Gentner, D., & Goldin-Meadow, S. (Eds.). (2003). *Language in mind: Advances in the study of language and thought*. Cambridge, MA: MIT Press.
Gleitman, L. R., Gleitman, H., Miller, C., & Ostrin, R. (1996). Similar, and similar concepts. *Cognition, 58*(3), 321–376.
Gleitman, L. R., January, D., Nappa, R., & Trueswell, J. C. (2007). On the give and take between event apprehension and utterance formulation. *Journal of Memory and Language, 57*(4), 544–569.
Gleitman, L., & Papafragou, A. (in press). New perspectives on language and thought. In K. Holyoak & R. Morrison (Eds.), *Cambridge handbook of thinking and reasoning*. New York: Cambridge University Press.
Hagen, J. W., & Kingsley, P. R. (1968). Labeling effects in short-term memory. *Child Development, 39*, 113–121.
Hermer-Vazquez, L., Spelke, E. S., & Katsnelson, A. S. (1999). Sources of flexibility in human cognition: Dual-task studies of space and language. *Cognitive Psychology, 39*(1), 3–36.
Hoffman, J. E., Landau, B., & Pagani, B. (2003). Spatial breakdown in spatial construction: Evidence from eye fixations in children with Williams syndrome. *Cognitive Psychology, 46*(3), 260–301.
Jackendoff, R. (1987). On beyond zebra: The relation of linguistic and visual information. *Cognition, 26*(2), 89–114.
Landau, B., Dessalegn, B., & Goldberg, A. M. (2011). Language and space: Momentary interactions. In V. Evans & P. Chilton (Eds.), *Language, cognition and space: The state of the art and new directions*. London: Equinox Publishing.
Landau, B., & Stecker, D. (1990). Objects and places: Geometric and syntactic representations in early lexical learning. *Cognitive Development, 5*(3), 287–312.
Levinson, S. C., Kita, S., Haun, D. B. M., & Rasch, B. r. H. (2002). Returning the tables: Language affects spatial reasoning. *Cognition, 84*(2), 155–188.
Li, P., & Gleitman, L. (2002). Turning the tables: Language and spatial reasoning. *Cognition, 83*(3), 265–294.
Loewenstein, J., & Gentner, D. (2005). Relational language and the development of relational mapping. *Cognitive Psychology, 50*, 315–353.
Lucy, J. A. & Gaskins, S. (2003). Interaction of language type and referent type in the development of nonverbal classification preferences. In D. Gentner & S. Goldin-Meadow (Eds.), *Language in mind: Advances in the study of language and thought* (pp. 465–492).
Lupyan, G., & Spivey, M. J. (2010). Redundant spoken labels facilitate perception of multiple items. *Attention, Perception, & Psychophysics, 72*(8), 2236–2253.
Lupyan, G. (2008). The conceptual grouping effect: Categories matter (and named categories matter more). *Cognition, 108*, 566–577.
Macnamara, J. (1978). *How can we talk about what we see?* Unpublished mimeo, Department of Psychology, McGill University, Montreal.
Miller, C. A. (1998). It takes two to tango: Understanding and acquiring symmetrical verbs. *Journal of Psycholinguistic Research, 27*(3), 385–411.
Miller, G. A. (1994). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 101*(2), 343–352.
Munnich, E., & Landau, B. (2003). The effects of spatial language on spatial representation: Setting some boundaries. In D. Gentner & S. Goldin-Meadow (Eds.). *Language in mind: Advances in the study of language and thought* (pp. 113–155). MIT Press.
Papafragou, A., Hulbert, J., & Trueswell, J. C. (2008). Does language guide event perception? Evidence from eye movements. *Cognition, 108*(1), 155–184.
Pederson, E., Danziger, E., Wilkins, D., Levinson, S. C., Kita, S., & Senft, G. (1998). Semantic typology and spatial conceptualization. *Language, 74*(3), 557–589.
Reese, H. W. (1975). Verbal effects in children's visual recognition memory. *Child Development, 46*(2), 400–407.
Roberson, D., & Davidoff, J. (2000). The 'Categorical Perception' of colours and facial expressions: The effect of verbal interference. *Memory and Cognition, 28*(6), 977–986.
Roberson, D., Davies, I., & Davidoff, J. (2000). Color categories are not universal: Replications and new evidence from a Stone Age culture. *Journal of Experimental Psychology: General, 129*(3), 369–398.
Rosch Heider, E., & Olivier, D. C. (1972). The structure of the color space in naming and memory or two languages. *Cognitive Psychology, 3*(2), 337–354.
Shusterman, A., Lee, S. A., & Spelke, E. S. (2011). Cognitive effects of language on human navigation. *Cognition, 120*(2), 186–201.
Smith, L. B., Jones, S. S., & Landau, B. (1996). Naming in young children: A dumb attentional mechanism? *Cognition, 60*(2), 143–171.
Soja, N. N., Carey, S., & Spelke, E. S. (1991). Ontological categories guide young children's inductions of word meaning: Object terms and substance terms. *Cognition, 38*, 179–211.
Spivey, M. J., Tyler, M. J., Eberhard, K. M., & Tanenhaus, M. K. (2001). Linguistically Mediatedvisual search. *Psychological Science, 12*(4), 282–286.
Stefurak, D. L., & Boynton, R. M. (1986). Independence of memory for categorically different colors and shapes. *Perception & Psychophysics, 39*(3), 164–174.
Subrahmanyam, K., Landau, B., & Gelman, R. (1999). Shape, material, and syntax: Interacting forces in children's learning in novel words for objects and substances. *Language and Cognitive Processes, 14*(3), 249–281.
Treisman, A. (1977). Focused attention in the perception and retrieval of multidimensional stimuli. *Perception & Psychophysics, 22*(1), 1–11.
Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12*(1), 97–136.
Treisman, A., & Schmidt, H. (1982). Illusory conjunctions in the perception of objects. *Cognitive Psychology, 14*(1), 107–141.
Whorf, B. L. (1956). In J. B. Carroll (Ed.), *Language, thought, and reality: Selected writings of Benjamin Lee Whorf*. Cambridge, Mass: MIT Press.
Yoshida, H., & Smith, L. B. (2003). Shifting Ontological Boundaries: How Japanese- and English-speaking children generalize names for animals and artifacts. *Developmental Science, 6*(1), 1–17.