NORTHWESTERN UNIVERSITY

The Effects of Perceptual Alignment and Linguistic Contrast on Preschoolers' Indirect Property-Word Learning

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Psychology

By

Ruxue Shao

EVANSTON, ILLINIOS

September 2019

Abstract

Young children can sometimes acquire new vocabulary words—even property terms—through indirect learning (e.g. Carey & Bartlett, 1978). We explore two factors that contribute to this ability—perceptual alignment and linguistic contrast. We propose that spontaneous comparison processes lead children to notice key commonalities and differences that facilitate indirect property word learning. More specifically, we hypothesize that children can quickly map a new property word to its referent when the target property is presented as an alignable difference between two highly similar and alignable objects. To test this, we revisited the Carey and Bartlett paradigm, varying the alignability of objects that 3-and 4-year-olds saw while hearing a novel color word, *chromium*. The first four studies focused on perceptual alignment. In Experiments 1 and 2, we found that children in the High Alignment condition were significantly better than those in the Low Alignment condition at identifying chromium objects in a subsequent task. Experiments 3 tested a direct pedagogy condition; as predicted, the results differed markedly from those of the indirect learning condition used in Experiments 1 and 2. Experiment 4 showed that the learning gained through this indirect paradigm was robust. Experiments 5 and 6 focused on alignment and contrast within the linguistic input. We found that children's indirect word learning was better when presented with high-quality linguistic cues that included semantic contrast in an alignable parallel syntactic structure (e.g., "the chromium one, not the blue one") than when presented with low-quality linguistic cues.

Across the six experiments, we found strong evidence that high perceptual alignment plays a pivotal role in children's indirect word learning. Linguistic cues also influenced children's indirect word learning, especially for the 4-year-olds. Overall, these studies shed light on ways of increasing referential transparency that can promote spontaneous indirect learning.

Keywords: indirect word learning; property words; perceptual alignment; linguistic contrast

Acknowledgements

The body of work presented here would not have been possible without the support and encouragement of my mentors and friends in the Northwestern University community and beyond. I am extremely thankful to my advisor, Dedre Gentner, for her inspiration and guidance throughout this six-year long journey. I am still at awe at her dedication, enthusiasm, and seemly endless energy for research. I would also like to thank my committee members, Sue Hespos and David Uttal, for their insight, suggestions, and support toward my dissertation as well as other lines of research. Lots of appreciation to the wonderful people in my department, who are not only my colleagues but also my friends: Nina Simms, Christian Hoyos, Francisco Maravilla, Erin Anderson, Sandy LaTourrette, and many, many more. Special thanks to Daikang Yan and Michelle Zhang for always being there for me through all the ups and downs of graduate school, for the late-night marathon phone calls and spontaneous parties, and the many bottles of wine that we have shared in celebration and consolation. Most of all, I am thankful for my parents, Yuanhua Shao and Na Li, and my brother, Scott Shao, for their love and support.

Table of Contents

Lists of Tables and Figures

Figure 15. (a) Mean number of chromium objects (max.= 3) correctly endorsed as *chromium* (hits); and (b) mean number of non-chromium objects (max. =5) incorrectly endorsed as *chromium* (false alarms) by 3-year-olds in Experiment 5's Meaning Assessment task.

Figure 16. (a) Mean number of chromium objects (max.= 3) correctly endorsed as *chromium* (hits); and (b) mean number of non-chromium objects (max. =5) incorrectly endorsed as *chromium* (false alarms) by 4-year-olds in Experiment 6's Meaning Assessment task.

Figure 17. (a) Mean number of chromium objects (max.= 3) correctly endorsed as *chromium* (hits); and (b) mean number of non-chromium objects (max. =5) incorrectly endorsed as *chromium* (false alarms) by 3-year-olds in Experiment 6's Meaning Assessment task .

Figure 18. Mean number of chromium objects (max.= 3) correctly endorsed as *chromium* (hits) by 4-year-olds in Experiments 1, 5, and 6.

Figure 19. Mean number of chromium objects (max.= 3) correctly endorsed as *chromium* (hits) by 3-year-olds in Experiments 1, 5, and 6.

Chapter 1: Introduction

When thinking about early word acquisition, people often envision a situation in which a teacher spontaneously provides the learner with a distinct object of interest and a clear label for the object. For example, a helpful adult may point to a rubber duck and tell a child, “Look! This is a duck!” Over the past decades, we have accumulated much knowledge about how children acquire noun labels in this kind of direct naming situation. Researchers have characterized factors that support early word learning, including perceptual salience (e.g. Landau, Smith, & Jones, 1988; Pruden, Hirsh-Pasek, Golinkoff, & Hennon, 2006; Yu & Smith, 2012), perceptual individuability (Gentner, 1982; Gentner & Boroditsky, 2001); conceptual knowledge (e.g. Booth & Waxman, 2002; Clark, 1987; Fernald, Zangl, Portillo, & Marchman, 2008; Markman, 1989), prosodic cues (e.g. Fernald & Mazzie, 1991; Jusczyk, Houston, & Newsome, 1999; Mattys, Jusczyk, Luce, & Morgan, 1999), joint attention (e.g. Brooks & Meltzoff, 2008; Tomasello & Farrar, 1986), social context (e.g. Akhtar & Tomasello, 2000; Baldwin, et al., 1996; Behne, Carpenter, & Tomasello, 2005; Ferguson & Lew-Williams, 2016; Tomasello & Barton, 1994), and syntactic cues (e.g. Bloom & Kelemen, 1995; Fisher, Gertner, Scott, & Yuan, 2010; Gleitman, 1990). Any or all of these cues can contribute to what Cartmill et al. (2013) called “referential transparency”—the degree to which the situation in which a word is uttered is conducive to the child’s learning the word.

In this work we investigate another source of referential transparency—available comparisons in the environment—in the context of indirect learning. Although most studies of word learning have involved direct word learning, children can also acquire new words through a wide range of indirect situations. Infants as young as 18-months of age can learn labels for novel objects through overhearing others’ conversations (e.g. Akhtar, 2005; Callanan, Akhtar, &

Sussman, 2014; Foster & Hund, 2012; Gampe, Liebal, & Tomasello, 2012). At around the same age, they can also learn words for objects through media consumption, especially with the guidance of an adult viewer (e.g. Krcmar, Grela, & Lin, 2007; Roseberry, Hirsh-Pasek, Parish-Morris, & Golinkoff, 2009). By preschool, children can also acquire new words through joint reading activities (e.g. Blewitt, Rump, Shealy, & Cook, 2009; Justice, 2002; Sénéchal, Thomas, & Monker, 1995).

Most studies on indirect word learning have focused on children's acquisition of concrete nouns—a reasonable first step, since children's early vocabulary is strongly noun-focused (Bornstein et al., 2004; Braginsky et al., 2015; Gentner, 1982, 2006; Gentner & Boroditsky, 2001, 2009). However, it leaves open the question of whether children can acquire other kinds of words—for example, property terms—through indirect measures, and if so, how they are able to do so. On the face of it, acquiring property words through indirect learning seems likely to be quite difficult.

Property words are conceptually demanding; because they denote specific properties of objects, a child must process the object referent before they can comprehend the target property. A child who hears "Look at the bifish dax" needs to be able to connect *dax* with the referent object, as well as to discover the intended dimension (e.g., color) and the intended value along the dimension (Gasser & Smith, 1998; Sandhofer & Smith, 1999). Another challenge is that there is considerable evidence that during early language learning, English-speaking children acquire a focus on nouns. Sandhofer and Smith (2007) found that children with fewer nouns in their productive vocabularies learned more adjectives than those with more nouns, unless strong syntactic cues were present. In addition, property words have relatively low frequency in parental input (Salerni, Assanelli, D'Odorico, & Rossi, 2007; Sandhofer, Smith, & Luo, 2000;

Tribushinina & Gillis, 2012).Not surprisingly, then, corpus studies have found that children are slower to learn property words than concrete nouns (e.g. Bates et al., 1994; Braginsky, Yurovsky, Marchman, & Frank, 2015; Bornstein et al., 2004; Gentner, 1982). The same pattern is found in experimental studies. For example, 14-month-olds who were directly taught a novel noun correctly extended the label to object categories, and not to same-property items (Booth & Waxman, 2003; Waxman, 1999; Waxman & Booth, 2001); but when taught a novel adjective, 14-month-olds, and even 18-month-olds, showed no preference between object-based vs. property-based extensions (Booth & Waxman, 2003; 2009; Waxman & Booth, 2001). Taylor and Gelman (1988) found similar patterns with 2-year-olds in a more complex word-learning task. Overall, research suggests that young children learn early to attach words to objects, and only slowly learn words for properties. Thus, even under direct presentation conditions, property terms are learned rather late compared to nouns.

Children can learn property words through indirect presentation

The above findings suggest that indirect learning of property terms should be extremely challenging. However, a few studies, including that of Carey and Bartlett (1978), have shown that children can also learn adjectives via indirect learning. Because Carey and Bartlett's study is central to our discussion, we describe it in some detail.

In Carey and Bartlett's (1978) classic study, 36-to 46-month-olds were led to connect a novel word (*chromium*) with an unfamiliar color (dark olive-green). During snack time at preschool, their teacher asked, "You see those two trays over there? Bring me the chromium one. Not the red one, the chromium one." (Carey & Bartlett, 1978, p.18). This was the full extent of the initial exposure: children's exposure to the new word and color was embedded within a task, without a direct referential label nor any pedagogical cues to signal the interaction as a potential learning experience. Thus, this paradigm serves as an apt method for studying indirect learning.

Seven-to-ten days after the initial exposure, children were assessed on their understanding of the word *chromium* and the color dark olive-green. They were exposed to the word and color again ten weeks later in two encounters that occurred two days apart. Seven-to-ten days after the second encounter, the same battery of tests was once again administered. The assessment included a color-naming task in which children were asked to label colors, including dark olive-green, and a comprehension task in which children were asked to pick *chromium* from an array of nine colors. Carey and Bartlett (1978) found that over half of the children retained knowledge of the word, even after this delay. Compared to an age-matched control group that did not get the exposure, those in the experimental condition were significantly better at the *chromium* comprehension task by the second round.

Not all children acquired the meaning of the word *chromium*—there was high variability within the experimental group, and only two children showed full command of the word. However, the Carey and Bartlett (1978) study is a milestone study for at least three reasons. First, it showed that children can learn a word's meaning from one exposure. This kind of rapid learning from one exposure (*fast mapping*) has inspired many further studies (e.g. Bion, Borovsky, & Fernald, 2013; Golinkoff, Hirsh-Pasek, Bailey, & Wenger, 1992; Horst & Samuelson, 2008; Spiegel & Halberda, 2011). Second, this research utilized indirect word learning, rather than direct labeling. Third, it focused on property terms, which, as reviewed above, are typically learned rather late.

This pattern of relatively slow learning of property words makes Carey and Bartlett's findings all the more remarkable. After all, in their task, children had no need to learn the meaning of *chromium*; they could satisfy the teacher's request simply by giving her the tray that was not red. Given the double challenge of learning property terms and learning via indirect input, the finding that some of their 3-to-4-year-old participants encoded and retained the new word's meaning is quite striking. How did children accomplish this feat?

Structure-mapping framework for indirect property-word learning

We propose that a key process for children's indirect property-word learning was difference detection through comparison. The two trays in Carey and Bartlett's (1978) study were extremely similar—identical in all but their color. As Gentner (2003, 2010) and colleagues have noted, comparison supports not only detecting commonalities but also detecting certain differences. In particular, there is abundant evidence that high similarity supports rapid, spontaneous difference detection (Gentner & Gunn, 2001; Gentner & Markman, 1994; Markman & Gentner, 1993; Sagi, Gentner, & Lovett, 2012). Gentner et al. (2015), using the structure-mapping framework, have suggested three reasons for this. First, high-similarity pairs invite spontaneous comparison; there's no need for a tutor to point out that the two things are alike (Gentner, Loewenstein, & Hung, 2006; Haryu, Imai, & Okada, 2011). Second, for high-similarity pairs, the process of structural alignment is fast and essentially effortless (Gentner & Hoyos, 2017; Kurtz & Gentner, 2013; Sagi, Gentner, & Lovett, 2012). Third, once structural alignment is achieved, *alignable differences*—that is, differences that play the same structural role in both items—tend to pop out as highly salient (Gentner & Markman, 1994; Gentner & Gunn, 2001; Markman & Gentner, 1993; Sagi, Gentner & Lovett, 2012). Thus, we hypothesize that in Carey and Bartlett's study, while simply complying with the request to "give me the chromium one, not the red one," children experienced a spontaneous comparison between the two highly similar trays, resulting in pop-out of the alignable difference in color—chromium vs.

red. The juxtaposition of this salient new color with the new word, *chromium* led some children to encode the connection, even without any intention to learn a new word.

Do linguistic cues matter for indirect property word learning?

In addition to highly alignable perceptual cues, children in the Carey and Bartlett study (1987) were given rich linguistic cues. They were told, “ Bring me the chromium one. Not the red one, the chromium one.” For experienced language users such as adults, this linguistic contrast is sufficient by itself to focus a listener to the distinguishing property of color, regardless of perceptual alignability (Au & Markman, 1987). However, there has been very little research on whether young children can profit from linguistic cues, and if so, which cues are important.

Heibeck and Markman (1987) varied the linguistic cues present in an indirect paradigm similar to that used by Carey and Bartlett (1978). They found that regardless of the informativeness of the linguistic cues provided (“Bring me the chartreuse one, not the red one.” or “Bring me the chartreuse one, not the other one.”), 2-to 4-year-olds learned novel property words. The researchers concluded that children mainly rely on non-linguistic cues when the linguistic and non-linguistic cues converge on the same information.

However, research by Au and colleagues (Au & Markman, 1987; Au, 1990; Au & Deframbiose, 1990) suggests that linguistic cues do promote property word learning under certain circumstances. They found that in direct word learning paradigms, children benefited from hearing a novel property word used in contrast with a known property word of the same dimension. However, this was only *when* the contrast was either consistent with the children’s previous knowledge and beliefs or emphasized through repetition. Although these findings shed some light on how linguistic cues affect property word learning, there are still many questions to be answered, especially when considering indirect learning situations.

Overview

In this paper, we present a series of studies examining how perceptual alignment and linguistic contrast contributes to referential transparency in indirect property word learning. We propose that high perceptual alignment invites spontaneous comparison and highlights the referent property. If this account is correct, then we should find that highly aligned pairs are important to young children's success in this paradigm. To test this, we adopted a paradigm similar to that used by Carey and Bartlett (1978) but varied the alignability of the pairs. Our account predicts that (1) when the object pairs are highly alignable, as in Carey and Bartlett's experiment, young children will be likely to learn the new color term (*chromium*) even with indirect exposure, and (2) this kind of spontaneous word learning should be much less likely when object pairs are less easy to align. In Experiments 1-4, we gave all children highly informative linguistic cues, as in Carey and Bartlett's (1978) study. In Experiments 5 and 6, we explored how different linguistic cues influence children's indirect property word learning, as well as how the informativeness of linguistic cues interacts with high versus low perceptual alignment.

Chapter 2: High Perceptual Alignment Promotes Indirect Property-Word Learning

In this chapter, we present two studies examining the role of perceptual alignment in indirect property word learning. In Experiment 1, we manipulated the alignability of the perceptual cues by showing children either highly similar and alignable object pairs, or object pairs that were low in overall similarity and alignability. In Experiment 2, we explored whether differences between high and low perceptual alignment were due to differences in the overall amount of information available. Since our goal here was to examine the unique role of perceptual alignment, we presented all children with the same highly informative linguistic cues.

Experiment 1

We presented 3-and 4-year-olds with a pair of objects—one blue, one dark olive green ("chromium"[1])—and asked the child to "point to the chromium one, not the blue one". There were two conditions: children in the High Alignment (HA) condition saw two objects that were highly similar (identical except for color) and therefore easy to align, whereas those in the Low Alignment (LA) condition saw objects that were less similar and therefore more difficult to align. After a break, we assessed their understanding of the word *chromium* through a yes-no classification task. We predicted that children in the HA condition would be more likely to encode the new chromium color and would therefore show more accurate classification than those in the LA condition.

Methods.

Participants. A total of 128 typically developing children participated in this study, including 65 3-year-olds (*M* = 42.89 months, *SD* = 2.89 months, females = 32) and 63 4-year-olds (*M* = 54.05 months, *SD* = 3.28 months, females = 32). Children were assigned to either the High Alignment (HA) or Low Alignment (LA) condition, with age and gender counterbalanced across the two conditions. An additional 25 children (five 4-year-old and twenty 3-year-olds) were tested but excluded from further analyses, six children due to failing to complete the experiment, four children due to experimenter error, and fifteen children for responding either all-yes or all-no in the Meaning Assessment task (see Methods section below). All children were recruited from the greater Chicago area through a voluntary participant pool or local preschools

[1] In the following text, chromium will be used to refer to the color and *chromium* will be used to refer to the word.

and the racial and economic composition of the sample reflected those of the local population (majority European-American, middle- and upper-middle-class). Children were given a small gift for participation.

Procedure and materials. The experiment took place in a single session. Children were initially exposed to the novel word, *chromium*, in an indirect learning context. After a break of approximately ten minutes, they were assessed on their understanding of the novel word through a yes-no classification task.

Initial Exposure. The Initial Exposure to the novel word, *chromium,* was embedded within a pointing task in which children were asked to point to one of two options. The task was designed to be fairly easy to preserve the sense of a fun game. For example, on the first trial, children saw a lion and a cow and were requested to "point to the lion—the lion, not the cow" or vice versa. The *chromium* exposure trial was always the third trial and followed by a fourth trial to prevent recency effects. Appendix A shows the full set of materials used in the pointing game. The first two trials and the final trial were identical for the HA and LA conditions.

On the *chromium* exposure trial, children in both the HA and LA conditions saw geometric shapes made of foam and of approximately the same size. Those in the HA condition saw two highly similar objects—a chromium square and a blue square; whereas those in the LA condition saw two objects that were similar in size but not shape—a chromium circle and a blue square (Figure 1). All children were asked, "Look at these two! Can you point to the chromium one? The chromium one, not the blue one." Left/right placement of the objects was counterbalanced in all trials. For the first two and final trials, the target object was also counterbalanced.

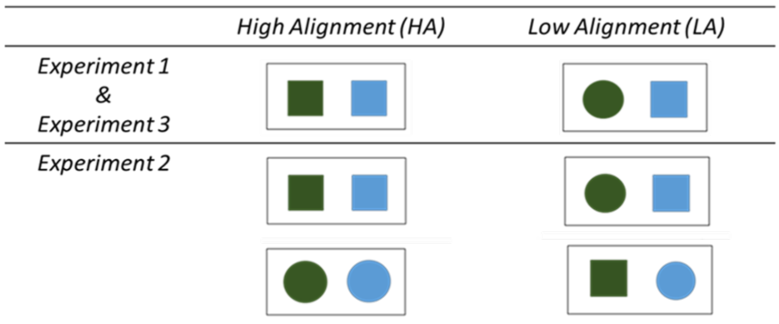


Figure 1. Initial Exposure objects presented to children in Experiments 1, 2, and 3.

Break. Between the Initial Exposure and Meaning Assessment, children completed a brief unrelated filler task for approximately 5 minutes (See Appendix B). After this task, children went on to a yes-no classification task intended to serve as a warm-up for the key Meaning Assessment task. In the warm-up task, children were given a series of 12 easy classification decisions: e.g., Child was asked “Is this an animal” while being shown a picture of a truck (See Appendix C). Children were shown a Yes Box with a smiley face and a No Box with a frowny face and were asked to put the cards in the corresponding box. No corrective feedback was given. To be included in the final analysis, children had to reach a predetermined criterion of 8 of 12 trials correct. No children were excluded in Experiment 1 due to failure to understand the task.

Meaning Assessment. The Meaning Assessment task immediately followed the warm-up classification task and was continuous with it. Children were shown eight geometric figures of various colors and shapes, one at a time in a pseudo-random order so that no consecutive objects shared the same color or shape (Figure 2). As each object was presented, the experimenter asked, “Look at this one! Is this a chromium one?” The object was placed in the Yes Box or the No Box based on the child’s answer. If the child failed to answer, or indicated that he/she did not know, the experimenter repeated the question. If the child still did not answer, the experimenter placed the Yes Box and No Box side by side in front of the child and asked, "Is this a chromium one? What do you think? Which box should we put it in?" No participant required more than three probes to answer a question.

The eight figures consisted of three chromium figures—a square, a circle, and a hexagon—and five non-chromium figures—a blue square, a blue circle, a maroon triangle, an orange circle, and a purple square. (See Figure 2a). Among the chromium figures, the chromium square had been seen in the Initial Exposure in the HA condition; likewise, the chromium circle had been seen in the LA condition. The chromium hexagon was new to both groups. In addition, the maroon triangle had been seen by all children in one of the filler trials of the Initial Exposure.

(a)

(b)

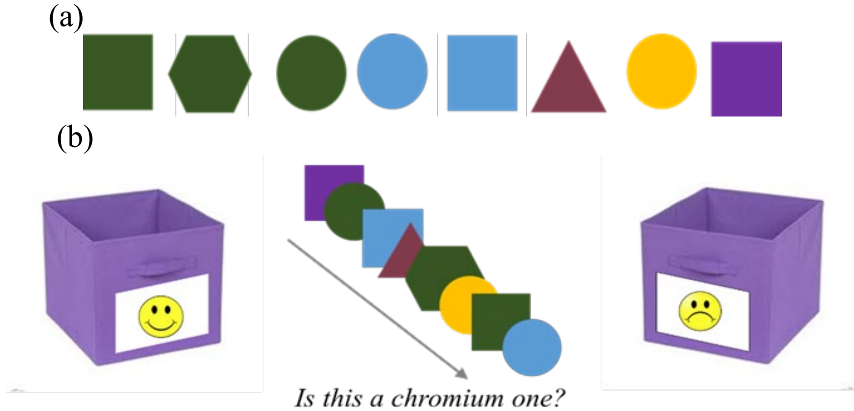


Figure 2. Schematics of (a) Meaning Assessment objects presented to children in Experiments 1 & 2; (b) Meaning Assessment Procedures. Objects were presented in pseudo random order in all studies.

Results.

Initial Exposure. Three-and four-year-olds in both the HA and LA conditions pointed to the target with high accuracy. Of the 128 children, 124 (96.9%) pointed to the chromium object on the first attempt. All four children who failed to do so were 3-year-olds, with three of them in the LA condition. These children were excluded from further analysis, since we do not expect children to map the word *chromium* to its corresponding color if they failed to identify the correct referent object.

Meaning Assessment. The Meaning Assessment task was used to assess whether children had learned the meaning of *chromium* during the Initial Exposure phase. We scored the number of hits (chromium objects correctly endorsed; max. = 3) and false alarms (non-chromium objects incorrectly endorsed; max. = 5) made by each child. As mentioned in the Participants section, we excluded children who responded either all-yes or all-no in the Meaning Assessment from further analyses. Eleven 3-year-olds and four 4-year-olds were excluded for this reason.

Correct identifications/hits. A two-way analysis of variance (ANOVA) with hits as the dependent variable and Age (3-year-olds or 4-year-olds) and Alignment (HA or LA) as the independent variables yielded a main effect of Alignment, $F(1,120) = 14.00$, $p < .001$, $\eta^2 = .10$, and a marginal effect of Age, $F(1,120) = 3.32$, $p = .07$, $\eta^2 = .03$, with 4-year-olds tending to make more hits than 3-year-olds ($M_{4yos} = 2.14$, $M_{3yos} = 1.80$). The interaction between Age and Alignment was non-significant, $F(1,120) = 0.05$, $p = .82$, $\eta^2 < .001$.

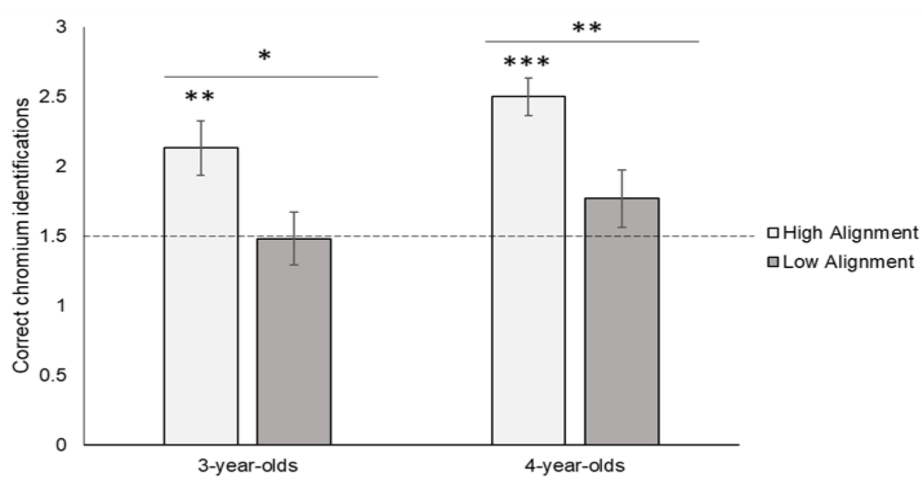


Figure 3. Mean number of chromium objects (max.= 3) correctly endorsed as *chromium* (hits) in Experiment 1's Meaning Assessment task. Error bars depict +/- one standard error. The dashed line represents chance performance. * $p < .05$, ** $p < .01$, *** $p < .005$.

Planned comparisons found that children in the HA condition made more hits than those in the LA condition, both among the 3-year-olds (M_{HA} = 2.13, SD_{HA} = 1.10; M_{LA} = 1.48, SD_{LA} = 1.02; $t(59)$ = 2.36, p = .01, d = 0.61, one-tailed *t*-test[2]) and among the 4-year-olds (M_{HA} = 2.50, SD_{HA} = .76; M_{LA} = 1.77, SD_{LA} = 1.15; $t(61)$ = 2.97, p = .002, d = 0.75, one-tailed *t*-test). In addition, only the HA groups performed better than chance (chance = 1.5 hits; both *p*s < .004, two-tailed *t*-tests[3]); the LA groups did not differ from chance (both *p*s > .19, two-tailed *t*-tests). Figure 3 shows the mean number of hits made by each Age × Alignment group.

We also found an Alignment effect when examining individual performance. First, we set a score of two or more hits (out of three) as the criterion for high performance[4]. In the HA condition, 23 out of 32 (71.9%) 3-year-olds and 27 out of 32 (84.4%) 4-year-olds performed above this criterion; whereas in the LA condition, 15 out of 29 (51.7%) 3-year-olds and 15 out of 31 (48.4%) 4-year-olds did so. Significantly more children reached criterion in the HA condition than in the LA condition for the 4-year-olds ($X^2(1, N = 63) = 9.18$, p = .002, φ = .38, chi-square test), but not for the 3-year-olds ($X^2(1, N = 61) = 2.63$, p = .11, φ = .21).

Memory check. To test for the possibility that the advantage of the HA groups over the LA groups was due to differential memory for the Initial Exposure object, we examined children's endorsement of the previously seen chromium object and new chromium objects separately. Among 4-year-olds, the number correctly endorsing the seen object did not differ

---

[2] We had predicted that children in the HA condition would outperform their age-matched peers in the LA condition. Thus, we used one-tailed *t*-tests throughout this paper when comparing the performances of children in the HA and LA conditions.
[3] We did not have directional hypotheses regarding children's performances compared to chance. Thus, we used two-tailed t-tests throughout this paper when comparing the performances of children in the HA and LA conditions.
[4] The probability of children correctly identifying all three chromium objects in the Meaning Assessment was higher than α = .05 (p(n = 3) = 0.125, binomial probability). Thus, we defined high performance as correctly endorsing at least two of the three chromium objects as *chromium*.

significantly between the HA and LA conditions (HA: 28 out of 32 [96.6%]; LA: 22 out of 31 [71.0%]), $X^2(1, N = 63) = 2.63$, $p = .11$, $\varphi = .21$). However, those in the HA condition were more likely to endorse new chromium objects ($M = 1.63$, $SD = 0.66$) than were those in the LA condition ($M = 1.06$, $SD = 0.89$), $t(61) = 2.84$, $p = .003$, $d = 0.73$, one-tailed *t*-test. Among 3-year-olds, there were non-significant trends for children in the HA condition to outperform those in the LA condition at endorsing both the seen object (HA: 24 out of 32 [75.0%]; LA: 15 out of 29 [51.7%]; $X^2(1, 61) = 3.57$, $p = .06$, $\varphi = .24$, chi-square test) and new chromium objects ($M_{HA} = 1.31$, $SD_{HA} = 0.78$; $M_{LA} = 0.97$, $SD_{LA} = 0.91$; $t(59) = 1.61$, $p = .06$, $d = 0.40$, one-tailed *t*-test). These findings suggest that the HA advantage in the Meaning Assessment task cannot be attributed simply to differential memory for the initial object.

Incorrect identifications/false alarms. Another possible explanation of the differences in hits is that children in the HA condition had a stronger tendency to respond positively in the Meaning Assessment task. We did not have enough items to conduct a *d'* analysis to examine children's sensitivity to chromium versus non-chromium objects. Instead, we examined false-alarm rates—i.e., rates of saying "yes" to non-chromium objects (max. = 5). Figure 4 shows the mean number of false alarms made by each Age × Alignment group.

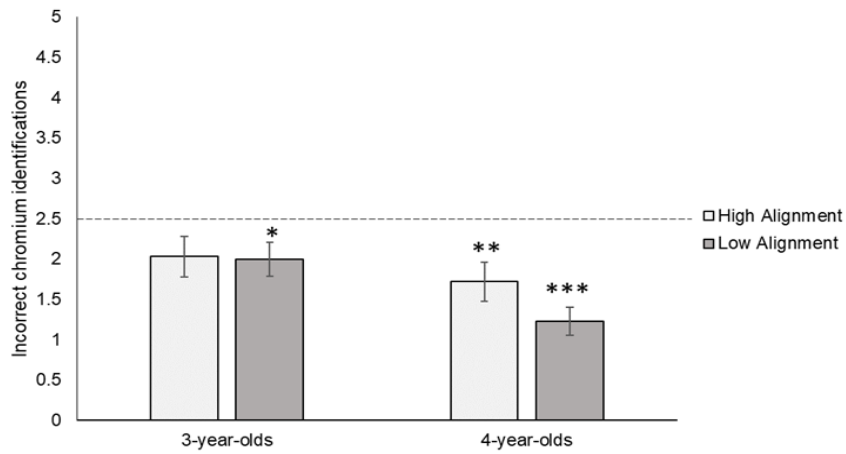


Figure 4. Mean number of non-chromium objects (max.= 5) incorrectly endorsed as *chromium* (false alarms) in Experiment 1's Meaning Assessment task. Error bars depict +/- one standard error. The dashed line represents chance performance. * $p < .05$, ** $p < .01$, *** $p < .005$.

We found that regardless of Age or Alignment condition, none of the groups made more false alarms than chance (chance = 2.5 false alarms). The mean number of false alarms made by 4-year-olds in the HA condition ($M = 1.72$, $SD = 1.35$), 4-year-olds in the LA condition ($M = 1.23$, $SD = 0.96$), and 3-year-olds in the LA condition ($M = 2.00$, $SD = 1.13$) were all significantly lower than chance (all *p*s < .03, two-tailed *t*-test). The mean number of false alarms made by 3-year-olds in the HA condition was not significantly different from chance ($M = 2.03$, $SD = 1.40$; $p = .07$, two-tailed *t*-test).

Experiment 1 Discussion.

As predicted, children in the HA group were significantly better than those in the LA group at mapping the word *chromium* to the color dark olive green. Although children in both conditions accurately pointed to "the chromium one, not the blue one" during the Initial Exposure task, only the HA group showed evidence of having extracted the meaning of *chromium* in a later assessment task. The LA groups—both the 3-year-olds and the 4-year-olds—performed at chance.

These findings are consistent with our proposal that perceptual alignability is instrumental in supporting indirect word learning. In both the experiment presented here and the original Carey and Bartlett (1978) study, children were exposed to the novel word and property in passing, in service of another task. Yet, even without any direct pedagogical cues, children who saw highly alignable objects encoded the meaning of *chromium*. The Low Alignment groups showed no such effect.

However, although the results so far are consistent with our proposal, they could also be explained in a very different way. Perhaps the performance differences between the HA and LA groups resulted simply from differences in the amount of available information. The two objects shown to the HA group differed in only one dimension – color. In contrast, the LA group saw objects that differed in *both* color and shape. From a hypothesis-testing point of view, the LA participants had to consider two possible hypotheses for the meaning of *chromium*—a color hypothesis (olive green) and a shape hypothesis (round)—whereas the HA participants only had to consider one—the color hypothesis. Thus, the HA advantage could have resulted from the children being able to rule out the erroneous hypotheses, rather than from their having experienced a spontaneous alignment process. Experiment 2 was designed to distinguish between these two possibilities.

Experiment 2

In Experiment 2 we equated the information that children received by giving all children two Initial Exposure trials. These trials were designed so that after the second trial, all children had sufficient information to logically conclude that color was the relevant dimension for the meaning of the word *chromium*. As in Experiment 1, we manipulated the alignability of the objects presented, with half of the children seeing easy-to-align pairs and half of the children seeing hard-to-align pairs. If children were engaged in hypothesis-testing or other types of intentional reasoning, then those in both the High Alignment (HA) and Low Alignment (LA) conditions would be able to correctly infer the meaning of *chromium*. However, if the HA advantage in Experiment 1 derived from spontaneous perceptual alignment, we should still see an advantage of HA over LA in Experiment 2's Meaning Assessment task.

Methods.

Participants. Since the 3-and 4-year-olds showed the same pattern of performance in Experiment 1, we examined only the younger age group in Experiment 2. A power analysis based on the findings of 3-year-olds in Experiment 1 suggested that at least 54 participants were needed to reach a power of .75 with an alpha of .05 (GPower 3.1, Faul, Erdfelder, Lang, & Buchner, 2007). We ran a total of 55 3-year-olds (M = 42.64 months, SD =2.80 months, females = 27). Children were randomly assigned to either the High Alignment (HA) or the Low Alignment (LA) condition, with approximately the same number of males and females in each condition. Six additional children participated in the experiment but were excluded due to either parental inference (one child), failing to complete the experiment (one child), failing to reach criterion for the yes-no sorting task used as a warm-up for the Meaning Assessment (one child), or answering all-yes on the Meaning Assessment task (three children[5]). Children were recruited through the same methods as in Experiment 1.

Materials and procedure. The materials and procedure for Experiment 2 were similar to those of Experiment 1 but with one crucial change: children received two Initial Exposure trials rather than one (See Figure 1, middle row). The key trials were embedded in the same simple

[5] Of the three children excluded for answering all-yes in the Meaning Assessment task, one was in the HA condition and two were in the LA condition.

pointing game as before. Children received a warm-up trial (e.g. "point to the lion—the lion, not the cow."), followed by the first *chromium* exposure trial. Then they received an easy trial (e.g. "point to the long one—the long one, not the short one."), followed by the second exposure trial and then a final easy trial. On both exposure trials, children were requested to "point to the chromium one—the chromium one, not the blue one". The order of the two *chromium* exposure trials and the order of the two easy trials were counterbalanced.

On the first exposure trial, the LA group saw two objects that differed in both shape and color, as in Experiment 1. For example, half of the LA group saw a chromium circle and a blue square. At this point, those children could have two possible hypotheses for the meaning of *chromium*: olive green or circle. On the second exposure trial, they saw a chromium square and a blue circle. Crucially, the target (non-blue) object on Trial 2 was of a different shape from the target object on Trial 1 but retained the same color (see Figure 1). Thus, by the end of the second trial, the LA group could rule out the shape hypothesis and arrive at the correct meaning of *chromium*—the olive-green color.

Children in the HA condition saw the same four objects as those in the LA condition. However, the objects were presented in two highly alignable pairs: e.g., a chromium circle and a blue circle on the first trial, and a chromium square and a blue square on the second. Thus, children in the HA condition did not receive additional information from the second trial.

Results.

Initial Exposure. As in Experiment 1, all children performed well in the Initial Exposure. Among the 55 children, 52 (94.5%) of them correctly pointed to the chromium object on both trials. The three children who did not do so were all in the LA condition and were excluded from further analyses.

Meaning Assessment. We compared the number of correctly identified chromium objects (hits, max. = 3) and incorrectly identified chromium objects (false alarms, max. = 5) that children made in the Meaning Assessment task.

Correct identifications/hits. Planned comparisons revealed that the HA group (M = 2.46, SD = 0.76) made significantly more hits than the LA group (M = 1.88, SD = 0.99), $t(50)$ = 2.35, p = .01, d = 0.66, one-tailed t-test. Children in the HA condition made significantly more hits than chance (chance = 1.5 hits), $t(25)$ = 6.45, $p < .001$, d = 1.26, two-tailed t-test. There was a non-significant trend for children in the LA condition to make more hits than chance, $t(25)$ = 1.98, p = .06, d = 0.39, two-tailed t-test. Figure 5a shows the mean number of hits made by children in the HA and LA groups in Experiment 2.

As in Experiment 1, to examine individual performance, we counted the number of children in each group who made at least two hits in the Meaning Assessment task. 18 out of 26 (69.2%) children in the HA condition and 10 out of 26 (38.5%) children in the LA condition reached the above criterion. A chi-squared test revealed that significantly more children reached criterion in the HA condition than in the LA condition ($X^2(1, N = 52) = 4.95$, p = .03, φ = .31).

Memory check. Since children had previously seen two of the three chromium objects presented in the Meaning Assessment, we compared how accurate children were for new versus seen chromium objects. There was no difference between the HA group's performance (M = 1.62, SD = 0.64) and the LA group's performance (M = 1.38, SD = 0.70) on the seen objects (the chromium circle and chromium square), $t(50)$ = 1.25, p = .11, d = 0.36, one-tailed t-test. However, the HA group was significantly more likely than the LA group to correctly identify the new object (the chromium hexagon) as *chromium*, (HA: 21 out of 26 [80.8%]; LA: 13 out of 26 [50.0%]), $X^2(2, N=52)=5.44$, p= .02, φ = .32.

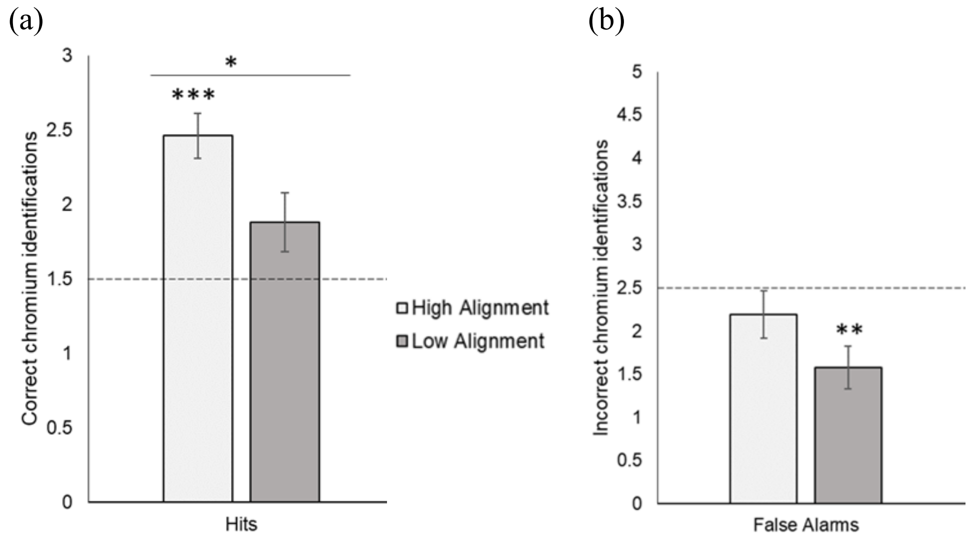


Figure 5. Mean number of (a) chromium objects (max.= 3) correctly endorsed as *chromium* (hits); and (b) non-chromium objects (max.= 5) incorrectly endorsed as *chromium* (false alarms) in Experiment 2's Meaning Assessment task. Error bars depict +/- one standard error. The dashed lines represent chance performance. * $p < .05$, ** $p < .01$, *** $p < .005$.

Incorrect identifications/false alarms. Children's high performance on hits was not due to an overall tendency to respond positively, as demonstrated by their low false alarm rates. Figure 5b shows the mean number of false alarms made by children in the HA and LA groups in Experiment 2. The mean number of false alarms made by the HA group ($M = 2.19$, $SD = 1.39$) did not differ from chance (chance = 2.5 false alarms), $p = .27$, two-tailed t-test. The mean number of false alarms made by the LA group ($M = 1.58$, $SD = 1.27$) was significantly lower than chance, $p = .001$, two-tailed t-test.

Experiment 2 Discussion

Experiment 2 replicated the high alignment advantage found in Experiment 1 and extended the findings to rule out an information-based explanation. Across two trials, children in both the HA and LA groups received adequate information to determine that *chromium* referred to a color, not a shape. As in Experiment 1, children in both groups correctly picked out the chromium object in the Initial Exposure; however, only the HA group was able to map the new word *chromium* to new instances in the subsequent Meaning Assessment task. Thus, the difference in performance in the two groups is not driven by unequal information. Instead, we suggest that the key distinguishing factor is a greater likelihood of spontaneous structural alignment by the HA group than the LA group.

In Experiment 2, there was a trend for children in LA condition to make more hits than expected by chance; whereas in Experiment 1, the mean number of hits made by children in the LA condition did not differ from chance. However, this difference appears due to differences in experimental design. In Experiment 1, only one of the three chromium objects in the Meaning Assessment task had been seen by children in the Initial Exposure. In Experiment 2, two of the three chromium objects had been seen. Across experiments, we find that children in both the HA and LA condition have good retention of previously seen chromium objects. However, if we only examine children's endorsement of the new chromium object, we still see an advantage of the HA groups over the LA groups in both Experiment 1 and Experiment 2.

Overall, our evidence indicates that the difference in performance between children who saw easy-to-align objects and those who saw hard-to-align objects is not driven by unequal information. Instead, we suggest that the key distinguishing factor is a greater likelihood of spontaneous structural alignment by the HA group than the LA group.

Chapter Discussion

How does alignability promote learning in such a context? Across two experiments, we found evidence suggesting that high perceptual alignability supports indirect property word learning, over and beyond the sheer amount of information presented. Following Gentner and Hoyos (2017), we suggest that high overall similarity of the objects benefitted the HA group in two ways. First, it prompted spontaneous comparison of the two objects. Second, the objects' high similarity enabled a fast and essentially effortless alignment, making the alignable difference of color "pop out" for easy detection (Sagi et al., 2012). Once the color of the target object was highlighted in this way, it was a conspicuous, ready candidate for the meaning of the word *chromium*, allowing children to link the new word to the salient color. Indeed, we suggest that this linkage occurred even without any deliberate attempt to learn the word's meaning—a point to which we return later.

The LA group (like the HA group) was highly accurate at pointing to the chromium object when asked to "point to the chromium one, not the blue one." However, (1) the disparity of the items made them unlikely to trigger spontaneous comparison, and (2) even if children had compared the objects, the lack of overall similarity would have made the alignment less fluent, and therefore less likely to result in pop out of the alignable difference. Of course, children could have engaged in deliberate comparison, as an adult might do, in order to derive the meaning of *chromium*. In this case, the LA group would have had two hypotheses as to the meaning of chromium after Trial 1 and could have selected the correct meaning (color) after Trial 2. But this did not happen. We saw no evidence that the LA group formed a link between the novel word *chromium* and its corresponding property. The results of Experiment 2 are consistent with our claim that the HA advantage stems from the greater likelihood of spontaneous comparison, and that this process did not require a deliberate intention to learn the meaning of chromium.

Chapter 3: Direct Versus Indirect Property-Word Learning

We have suggested that success in this task did not require any conscious intention to learn the meaning of the word *chromium*. Rather, both groups of children may simply have wanted to comply with the experimenter's request to point to the object that was "not the blue one." Given a readily alignable pair, we propose that a spontaneous comparison occurred, rendering the 'chromium' color more salient (as an alignable difference). The HA group was thus able to link the new word *chromium* with the new color, without any prior intention of doing so.

In order to test this claim of spontaneous implicit alignment, Experiments 1 and 2 were designed to minimize any demand to learn a new word. We exposed children to the novel word and target property in a communicative exchange that was designed to seem like a pointing game. Further, an unrelated filler task intervened between the initial exposure and the later assessment. Finally, we presented the assessment of the novel word as one of many episodes in a color sorting task. However, despite our best efforts to create an incidental learning situation, it is still possible that children were intentionally trying to learn the meaning of *chromium*. If so, this would call for a different kind of processing explanation than the one we have proposed.

Thus, in Experiment 3, we bore down on the assumption that the task is functioning as an implicit learning situation by comparing it to a direct-learning version of the task. In this study, children were directly instructed to learn the word *chromium*. This enables us to test between two possibilities. One possibility is that children in our prior studies were trying to learn the word, and this happened to be easier with high perceptual alignability. In this case, asking children to learn the word should not greatly alter their performance. The second possibility, and the one we propose, is that children in our prior studies were simply playing a pointing game, and the high-alignment advantage arose through spontaneous implicit comparison. In this case, instructions to learn the word should improve the performance of the low-alignment group (which lacks the advantage of spontaneous comparison) but should have little or no effect on the performance of the high-alignment group.

Experiment 3

Methods.

Participants. A power analysis based on the findings of 4-year-olds in Experiment 1 suggested that at least 40 participants were needed to reach a power of .75 with an alpha of .05 (GPower 3.1, Faul et al., 2007). Forty-four 4-year-olds were included in Experiment 3 (M = 53.17 months, SD =2.65 months, females = 20). An additional nine children participated but were excluded: one child for failing to complete the experiment, two children due to experimenter error, two children for failing to reach criterion for the yes-no sorting task used as a warm-up for the Meaning Assessment[6], and four children for answering yes to all trials in the Meaning Assessment[7]. All children were recruited through the same measures as in previous experiments. Children were randomly assigned to either the HA or LA condition, with approximately equal numbers of males and females in each condition.

Materials and procedure. The materials and procedure for Experiment 3 were similar to those of Experiment 1, except that children were explicitly told that their goal was to learn the meaning of a new word. Before the Initial Exposure, the experimenter told the child, "We are

[6] These children did not reach the predetermine criterion (at least eight out of twelve correct) on the yes-no sorting task used as a warm-up for the Meaning Assessment task. See Chapter 1 for more details.
[7] Two of the children who answered yes to all trials were in the HA condition and two were in the LA condition.

going to play a fun game together! In this game, you have special mission and that is to learn a new word –*chromium*. Can you say *chromium*?" After the child repeated the word, the experimenter said, "Great! Remember, you need to figure out what *chromium* means by the end of the game, okay?" After the introduction, the experiment was conducted as in Experiment 1. At the end of the study, the experimenter asked, "In this game you were supposed to learn a new word. Do you remember what word it was?" If the child produced the word, the experimenter replied, "Right, the word was *chromium*! What do you think *chromium* means?" If the child indicated that they forgot or simply remained silent, the experimenter asked, "Was the word… *avocado*? Was it…*stegosaurus*? Was it…*chromium*?" If the child picked the correct word, the experimenter said, "Right, the word was *chromium*. What do you think *chromium* means?" If the child picked the wrong word, the experimenter replied, "Oh, I think that the word was *chromium*. What do you think *chromium* means?"

Results.

Initial Exposure. Regardless of condition, all children (44/44) pointed to the correct chromium object during the Initial Exposure.

Meaning Assessment. We next compared the number of correctly identified chromium objects (hits, max. = 3) and incorrectly identified chromium objects (false alarms; max. = 5) that children made in the Meaning Assessment task.

Correct identifications/hits. One-tailed *t*-tests revealed that, in contrast to the prior studies, the two groups did not differ in their number of hits (M_{HA} = 2.59, SD_{HA} = 0.73; M_{LA} = 2.32, SD_{LA} = 0.89; $t(42) = 1.11$, $p = .14$, $d = 0.33$, one-tailed *t*-test). Both groups made hits at significantly above chance rates (chance = 1.5 hits; HA: $t(21) = 6.97$, $p < .001$, $d = 1.49$, two-tailed *t*-test; LA: $t(21) = 4.29$, $p < .001$, $d = 0.92$, two-tailed *t*-test). In addition, there was no difference between the HA and LA conditions in the number of children who made two or more hits, $p = .26$, chi-square tests. 19 out of 22 (86.4%) children in the HA condition and 16 out of 22 (72.2%) children in the LA condition reached this criterion. Figure 6 shows the mean number of hits made by children in the HA and LA conditions in Experiment 1 and Experiment 3.

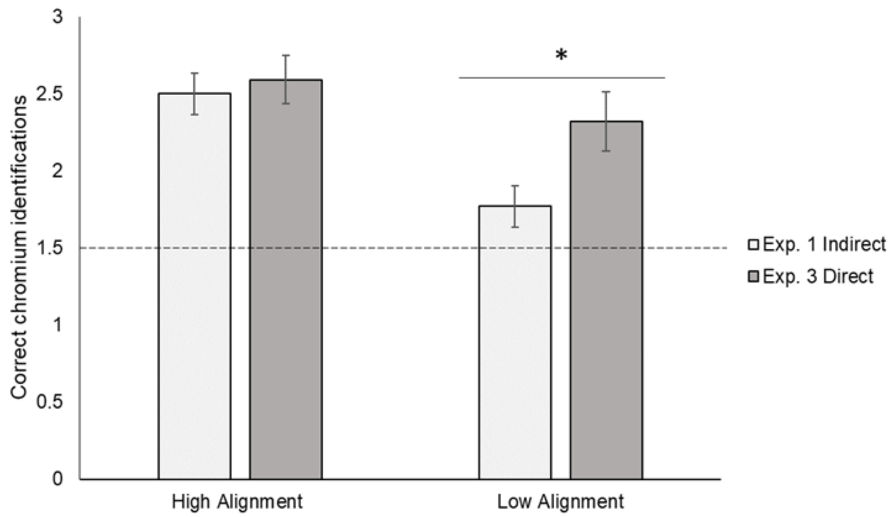


Figure 6. Mean number of chromium objects (max.= 3) correctly endorsed as *chromium* (hits) by children in Experiment 1 (Indirect) and Experiment 3 (Direct). Error bars depict +/- one standard error. The dashed line represents chance performance. * $p < .05$, ** $p < .01$, *** $p < .005$.

Memory check. The number of children to correctly endorse the chromium object that they had previously seen in the Initial Exposure did not differ between the HA condition (19 out of 22 [86.4%]) and the LA condition (19 out of 22 [86.4%]). There was a non-significant trend for children in the HA condition ($M = 1.73$, $SD = 0.55$) to correctly endorse more new objects than those in the LA condition ($M = 1.45$, $SD = 0.74$), $t(42) = 1.39$, $p = .09$, $d = 0.43$, one-tailed t-test.

Incorrect identifications/false Alarms. Both the HA and LA groups made significantly fewer false alarms than expected by chance (chance = 2.5 false alarms; $M_{HA} = 1.82$, $SD_{HA} = 1.14$; $M_{LA} = 1.59$, $SD_{LA} = 1.26$; both ps < .02, two-tailed t-tests), suggesting that their high hit rates were not due to an overall bias to respond positively—children discriminated between the chromium and non-chromium objects.

Cross-study comparison. To compare the effects of direct and indirect conditions, we compared the results of Experiment 3 with those of Experiment 1. The two studies were identical except for the early instruction to learn the new word *chromium* in Experiment 3. We first examined the number of chromium objects correctly identified in the Meaning Assessment task. A two-way ANOVA revealed a significant effect of Alignment (HA vs. LA), $F(1,103) = 7.78$, $p = .006$, $\eta^2 = .07$. There was a non-significant trend for children to perform better in Experiment 3 (direct) than in Experiment 1 (indirect), $F(1,103) = 3.14$, $p = .08$, $\eta^2 = .03$. We did not find an interaction between Alignment and Experiment, $F(1,103) = 1.60$, $p = .21$, $\eta^2 = .02$. There were no differences in false alarms, $F(3,103) = 1.37$, $p = .26$.

To test our prediction that direct instruction should preferentially benefit the LA group, we next conducted planned comparisons comparing the HA and LA groups in Experiment 3 (direct task) with those in Experiment 1 (indirect task). As predicted, for the HA groups, there was no difference between direct and indirect tasks in either hits or false alarms, $t(52) = .44$, $p = .33$, $d = 0.12$, and $t(52) = .28$, $p = .39$, $d = 0.08$, one-tailed *t*-tests. When presented with highly alignable objects, whether children received direct instruction or not did not influence their performance on the Meaning Assessment task. However, for the LA groups, children made a significantly greater number of hits in the direct condition (Experiment 3: $M = 2.32$, $SD = 0.89$) than in the indirect condition (Experiment 1: $M = 1.77$, $SD = 1.15$), $t(51) = 1.86$, $p = .03$, $d = 0.52$, one-tailed *t*-test. The number of false alarms made by the LA groups did not differ between the two experiments, $t(51) = 1.20$, $p = .12$, $d = 0.33$, one-tailed *t*-test.

Word memory check. We examined how well children were able to remember the word learning instructions that they were given at the beginning of the experiment. Of the 44 children included in the final analysis, 31 of them (70%, p = .01, binomial test) spontaneously produced some form of the word *chromium* when asked "what was the word that you were supposed to learn" (e.g. "chromium", "chromy"). Of the remaining 13 children, all but one[8] picked out the correct word from three possible options. When asked to say what *chromium* means, 17 children provided a color meaning (e.g. "black", "dark"), 4 children provided a shape meaning, (e.g. "circle", "shapes"), 2 children provided a texture meaning (e.g. "hard", "a little soft"), 12 children indicated that they did not know, and 9 children responded in other ways (e.g. "grill", "animals", "we're done with the game"). The number of children who gave a color meaning did not differ between the HA condition (9 out of 22 [40.1%]) and LA condition (8 out of 22 [36.4%]), ($X^2(1, N = 44) = .10, p = .76, \varphi = .47$).

Experiment 3 Discussion.

As predicted, word learning in the LA condition benefitted from direct instructions to learn. Under indirect learning conditions (Experiments 1 and 2), the number of children in the LA groups who reached criterion in the Meaning Assessment task did not differ from chance. But when given direct instructions to learn the word *chromium*, significantly more LA children reached criterion than expected by chance. Further, in contrast to prior studies, there was no significant difference between the HA and LA groups in performance on the Meaning Assessment task. The finding that LA children succeeded when asked to learn the word but failed to learn the word in the prior studies, suggests that children in our prior studies were not approaching the experiment as a word-learning task.

---

[8] This child was a female in the HA condition.

Unlike the LA group, the HA group was not affected by the addition of explicit instructions—children performed equally well on the Meaning Assessment task regardless of whether they were told that they had to learn a new word. This is consistent with the idea that children in the HA condition were learning via spontaneous alignment processes, rather than by intentionally seeking to learn the word *chromium*. The contrast between the LA group and the HA group (whose performance was unaffected by instruction) highlights the importance of perceptual alignment in promoting spontaneous, non-intentional word learning.

At the same time, the findings for the LA group show that perceptual alignment is not *necessary* for learning property words. The difference between the LA and HA groups found in the prior studies disappeared when children were given direct task instructions—evidence that pedagogical support can compensate for a lack of ideal perceptual conditions. Of course, this lack of difference could be due to a ceiling effect in our studies. With more complex stimuli, we might see an advantage of high-alignability even with direct instruction.

Taken together, the three studies so far provide evidence that perceptual alignment can support implicit learning of property terms. But how robust is this learning? Is it a fleeting impression, or will children retain these meanings over time? And is the learning narrowly construed, or can children apply the *chromium* concept to very different materials? In our experiments so far, we have assessed children's understanding of *chromium* ten minutes after the Initial Exposure, using objects similar to those seen in the Initial Exposure. In the next chapter, we ask whether children will retain the meaning of *chromium* over a delay, and whether they can apply it to a new set of objects.

Chapter 4: Robustness and Transfer in Indirect Property-Word Learning

In this chapter, we examined the strength of children's learning by (1) assessing children's understanding of the novel word's meaning after a longer delay; and (2) increasing the diversity of the transfer items in the Meaning Assessment Task. As in Experiment 2, both the high alignment and low alignment groups saw two exposure trials during the Initial Exposure, designed to provide sufficient information to allow children to reject a shape hypothesis and conclude that *chromium* refers to a color. As before, we assessed participants' knowledge of the word *chromium* ten minutes after the first exposure (Day 1). In addition, to investigate robustness under delay, we also assessed their understanding two-to-four days later (Day 2). Day 2 featured two tasks used to assess children's understanding of *chromium*: a Retention Task to assess the durability of the word-meaning mapping and a Transfer Task to assess whether children could extend the word to new kinds of materials. One further change was that, in the interest of generalizability, we developed new stimuli for both the Initial Exposure and Meaning Assessment.

Experiment 4

Participants.

A power analysis based on the findings of 4-year-olds in Experiment 1 suggested that at least 40 participants were needed to reach a power of .75 with an alpha of .05 (GPower 3.1, Faul et al., 2007). We ran a total of 41 4-year-olds (M= 53.54 months, SD= 3.61 months, females= 26) in Experiment 4. Participants were randomly assigned to either the High Alignment (HA) condition (n = 20, female = 12) or the Low Alignment (LA) condition, with approximately equal numbers of males and females in each group. An additional four 4-year-olds was tested but excluded from the analyses, one for failing to complete the task and three for teacher interference. All children were recruited through the same methods as previous experiments.

A group of 3-year-olds also participated in Experiment 4; however, they performed at chance in both the HA and LA conditions on Day 1, making it difficult to interpret their subsequent performance. We present their results separately later in the chapter.

Day 1 Materials and Procedure.

The procedure for Day 1 was the same as that of Experiment 2. Children were exposed to the novel word and color during a pointing task and assessed on their learning of the word-color mapping approximately ten minutes later through a yes-no sorting task. However, we used new stimuli for both the Initial Exposure and Meaning Assessment. During the Initial Exposure (Figure 7a), children in the HA condition saw two crosses (one chromium and one blue) and two upside-down triangles (also one chromium and one blue). Children in the LA condition saw a chromium cross paired with a blue circle, and a chromium upside-down triangle paired with a blue square. During the Meaning Assessment task (Figure 7b), children saw a set of 14 geometric shapes (as opposed to only 8 shapes in Experiments 1-3) created under the same guidelines as in previous studies. The set included five chromium objects, four blue objects and five objects of various colors. As in previous experiments, the set included all the objects presented across the Initial Exposure trials. In addition, the set also included objects with seen shapes and new colors (e.g. the orange cross), objects with new shapes and seen colors (e.g. the chromium hexagon), completely new objects (e.g. the yellow L shape), and a maroon triangle that had been seen by all children in one of the Initial Exposure's filler trials.

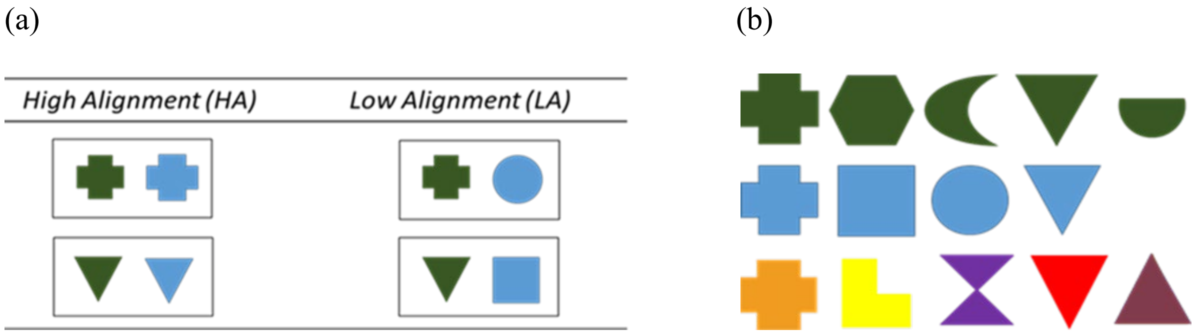


Figure 7. (a) Experiment 4 Day 1 Initial Exposure objects; (b) Experiment 4 Day 1 Meaning Assessment task and Experiment 4 Day 2 Retention Task objects.

Day 1 Results for 4-year-olds.

Initial Exposure. As in prior studies, children performed very well on the Initial Exposure task. 40 of the 41 (97.6%) children successfully pointed to the chromium object on both Initial Exposure trials. Only one 4-year-old (HA) was incorrect on the first key exposure trial. This child was excluded from further analyses.

Meaning Assessment.

Correct identifications/hits. As in Experiment 1, children in the HA group correctly endorsed more chromium objects than those in the LA group ($M_{HA} = 4.63$, $SD_{HA} = 0.96$; $M_{LA} = 3.71$, $SD_{LA} = 1.65$), $t(38) = 2.12$, $p = .02$, $d = 0.64$, one-tailed *t*-test. However, unlike Experiment 1—in which only children in the HA condition performed above chance—here, children made significantly more hits than expected by chance (chance = 2.5 hits) regardless of Alignment condition (both *p*s < .004, two-tailed *t*-tests). Figure 8 shows the mean number of hits made by 4-year-olds in Experiment 4.

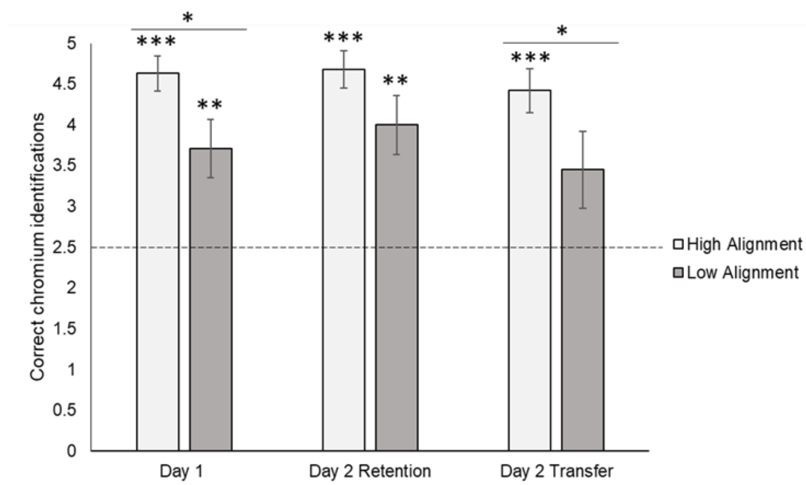


Figure 8. Mean number of chromium objects (max.= 5) correctly endorsed as *chromium* (hits) by 4-year-olds in Experiment 4 Day 1 Meaning Assessment task, Day 2 Retention task, and Day 3 Transfer task. Error bars depict +/- one standard error. The dashed line represents chance performance. * $p < .05$, ** $p < .01$, *** $p < .005$.

Next, we examined performance at the individual level. Because Experiment 4's Meaning Assessment task contained more objects than in previous studies, we were able to use performance above chance as the criterion for success for both hits and false alarms. A child needed to correctly identify all five chromium objects to perform significantly better than chance on hits ($p(n \geq 5) = 0.03$, binomial probability). In the HA condition, 15 out of 19 children (78.9%) did so, compared to 12 out of 21 (57.1%) children in the LA condition. There was no difference between the number of children scoring above chance in the two conditions, $X^2(1, N = 40) = 2.16$, $p = .14$, $\varphi = .23$, chi-square test.

Memory check. To examine if children's high hit rates were mainly driven by their memory for previously seen objects, we compared children's endorsement rates for the two seen objects and the three new chromium objects separately. Children in the HA condition performed better than those in the LA condition at identifying both previously seen chromium objects (M_{HA}

= 1.89, SD_{HA} = 0.32, M_{LA} = 1.57, SD_{LA} = 0.60; $t(38) = 2.11$, $p = .02$, $d = .66$, one-tailed *t*-test) and new chromium objects (M_{HA} = 2.74, SD_{HA} = 0.73, M_{LA} = 2.10, SD_{LA} = 1.22; $t(38) = 1.99$, $p = .03$, $d = 0.63$, one-tailed *t*-test). For both seen and new objects, all children performed significantly better than chance[9] (all *p*s < .04, two-tailed *t*-tests).

Incorrect identifications/false alarms. Children in the both the HA condition ($M = 1.32$, $SD = 1.57$) and LA condition ($M = 1.76$, $SD = 2.39$) made significantly fewer false alarms than expected by chance (chance = 4.5 false alarms; both *p*s < .001, two-tailed *t*-tests).

d prime. With the extended set of objects in the Meaning Assessment, we were able to examine how sensitive children were to the meaning of *chromium.* Combining hits and false alarms, we calculated the *d*' score for each individual child and compared the mean *d'* score of the HA group and LA group. There was a non-significant trend for the mean *d'* of 4-year-olds in the HA condition ($M = 2.20$, $SD = 0.86$) to be higher than that of those in the LA condition ($M = 1.57$, $SD = 1.46$), $p = .06$, one-tailed *t*-test. Figure 9 shows the mean *d'* scores made by 4-year-olds in the HA and LA conditions.

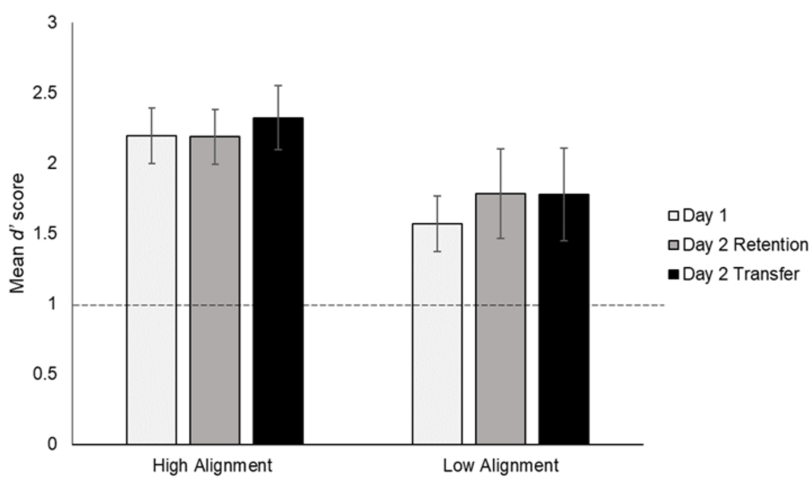


Figure 9. Mean *d'* scores by 4-year-old in Experiment 4. Error bars depict +/2 one standard error. The dashed line represents a *d*' score of 1.00.

[9] Chance rate for seen chromium objects was 1 hit; chance rate for new chromium objects was 1.5 hits.

Day 2 Materials and Procedure.

Two-to-four days (*M*= 2.43, *SD*= 0.75) after Day 1, children returned to the lab for Day 2 of the experiment. All children went through the same procedure: a Retention Task, a Reminder, and a Transfer Task.

The Retention Task was an exact replica of the Meaning Assessment Task on Day 1 and was used to assess how well participants retained the word meaning. After the Retention Task, participants received a refresher prompt, analogous to the Low Alignment Initial Exposure on Day 1. They were shown a card depicting a chromium fish, together with another card depicting a different looking blue fish. (Figure 10a) The experimenter asked, “Look at these! Can you point to the chromium one? The chromium one, not the blue one!” Crucially, children in the HA and LA conditions received the same Reminder. Afterwards, children completed the Transfer Task. This task used the same format as the Meaning Assessment task (and Retention task), except that children were shown 14 picture cards of differently colored and shaped fish. As shown in Figure 10b, five of the fish were chromium, four were blue, and the rest were of various colors. All fish in the Transfer task were new; neither of the two fish in the Reminder were shown again in the Transfer task.

(a) (b)

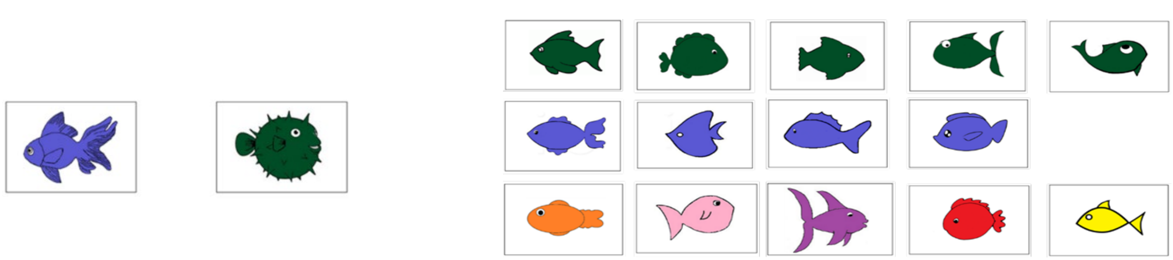


Figure 10. Stimuli presented in (a) Experiment 3 Day 2 Reminder; (b) Experiment 3 Day 2 Transfer Task objects.

Day 2 Results for 4-year-olds.

Retention task. Of the 40 4-year-olds included in Day 1's Meaning Assessment analyses, one child in the LA condition was excluded from the following analyses due to answering yes to all the trials in the Retention task.

Correct identifications/hits. Among the remaining 39 children, there was a non-significant trend for the mean number of hits (max. = 5) made by the HA group ($M = 4.68$, $SD = 1.00$) to be greater than that made by the LA group ($M = 4.00$, $SD = 1.62$), $t(37) = 1.57$, $p = .06$, $d = 0.50$, one-tailed t-test. Children in both conditions made significantly more hits than expected by chance (chance = 2.5 hits, both ps < .007, two-tailed t-tests). Figure 8 shows the mean number of hits made by 4-year-olds in the Retention Task.

At the individual level, 17 out of 19 (89.5%) children in the HA condition and 13 out of 20 (65.0%) children in the LA condition made significantly more hits than expected by chance (i.e., identifying all five chromium objects). There was a non-significant trend for more children to score above chance in the HA condition than the LA condition ($X^2(1, N = 39) = 3.29$, $p = .07$, $\varphi = .29$, chi-square test).

Incorrect identifications/false alarms. The mean numbers of false alarms made by children in both the HA condition ($M = 1.42$, $SD = 1.84$) and LA condition ($M = 1.55$, $SD = 2.44$) were significantly lower than expected by chance (chance = 4.5 false alarms, both ps < .001, two-tailed t-tests).

d prime. Combining hits and false alarms, we found that 4-year-olds in both the HA and LA conditions had high *d'* scores. We found no difference between the mean *d'* score of children in the HA condition ($M = 2.19$, $SD = 0.84$) and those in the LA condition ($M = 1.79$, $SD = 1.43$),

$p = .15$, one-tailed *t*-test. Figure 9 shows the mean *d'* scores made by 4-year-olds in the HA and LA conditions.

Reminder. When shown two fish and asked to "point to the chromium one, not the blue one," all 4-year-olds included in Day 1's Meaning Assessment analyses (40/40) chose the correct chromium object on the first attempt, consistent with previous results on the Initial Exposure Task.

Transfer task. As before, we excluded children who answered yes to all trials in the Transfer task ($n = 1$, LA condition, same child excluded in the Retention task). However, we included children who answered no to all trials in the Transfer task ($n = 4$, LA = 3), since an all-no response could indicate failure to transfer[10].

Correct identifications/hits. Among the remaining 39 children, those in the HA group made significantly more hits than those the LA group (M_{HA}= 4.42, SD_{HA}= 1.17; M_{LA}= 3.45, SD_{LA}= 2.11), $t(37) = 1.76$, $p = .04$, $d = 0.59$, one-tailed *t*-test. Figure 8 shows the mean number of hits made by 4-year-olds in Experiment 4. Children in the HA condition made significantly more hits than chance (chance = 2.5 hits), $t(18) = 7.16$, $p < .001$, $d = 1.64$, two-tailed *t*- test. There was a non-significant trend for children in the LA condition to make more hits than chance, $t(19) = 2.01$, $p = .06$, $d = 0.45$, two-tailed *t*-test. The number of children who correctly identified all five chromium objects did not differ between the HA condition (12 out of 19 [63.2%]) and the LA condition (12 out of 20 [60.0%]), $p = .84$, chi-square test.

---

[10] An analysis excluding the children who answered all-no in the Transfer task found a non-significant trend for the HA group ($M = 4.67$, $SD = .49$) to make more hits than the LA group ($M = 4.11$, $SD = 1.61$), p = .08, one-tailed t-test. The two groups did not differ in their mean number of false alarms, p = . 22, one-tailed *t*-test.

Incorrect identifications/false alarms. The mean numbers of false alarms made by children in both the HA and LA conditions (M_{HA} = 1.11, SD_{HA} = 2.26; M_{LA} = 1.35, SD_{LA} = 2.48) were significantly lower than chance (chance = 4.5 false alarms), both *p*s < .001, two-tailed *t*-tests.

d prime. Four-year-olds in both the HA and LA conditions showed high sensitivity to the novel word and color in the Transfer task, as evidenced by their high *d'* scores . There was a non-significant trend for children in the HA condition (*M* = 2.22, *SD* = 1.04) to have a higher mean *d'* score than those in the LA condition (*M* = 1.61, *SD* = 1.42), *t*(37) = 1.53, *p* = .15, *d* = .49, one-tailed *t*-test. Figure 9 shows the mean *d'* scores made by 4-year-olds in the HA and LA conditions.

Comparison of all tasks.

To compare overall performance in Experiment 4, we conducted a repeated-measures ANOVA with the 39 children who were included in all three tasks. We used Hits as the dependent variable, Task (Day 1 Meaning Assessment, Day 2 Retention, and Day 2 Transfer) as a within-subject independent variable and Alignment (HA and LA) as a between-subject independent variable. We found that the HA group performed significantly better than the LA group overall across the three tasks, *F*(1,37) = 5.02, *p* = .03. There was no significant main effect of Task (*p* = .23), nor an interaction between Task and Condition (*p* = .78).

Discussion.

In Experiment 4, we examined the robustness of incidental word learning through spontaneous alignment—first, by assessing children's understanding of the novel word after a delay of two-to-four-days; and, second, by assessing their ability to transfer the meaning of *chromium* to completely new objects. Four-year-olds performed slightly better than in

Experiment 1, in that both the HA and LA groups were above chance in number of correct identifications (hits) on the Meaning Assessment task, Retention task, and Transfer task (we will return to this point in the Discussion of 3-year-olds' Performance). However, the HA group performed better overall than the LA group: on Day 1, when tested shortly after the Initial Exposure, 4-year-olds in the HA condition scored significantly higher than did those in the LA condition. The high-alignment advantage also held on Day 2. The HA group scored higher than the LA group on the Transfer task, with a trend in this direction on the Retention task.

These results indicate that the insights from high-alignment comparison go beyond a momentary advantage. Having an alignable comparison in the environment can confer lasting benefits and can inform understanding of new situations.

Of course, seeing a benefit of high-alignment comparison over time is dependent on individuals actually forming insights during the initial exposure. As mentioned in the beginning of the chapter, in addition to the 4-year-olds presented above, we also ran 41 3-year-olds in Experiment 4. However, 3-year-olds in the HA condition did not show evidence of having learned the meaning of *chromium* in the Meaning Assessment task on Day 1. In the remaining part of this chapter, we briefly present the 3-year-old findings.

Performance of 3-year-olds in Experiment 4

Day 1 Results.

Initial Exposure. All 3-year-old children (100%) successfully pointed to the chromium object on both Initial Exposure trials.

Meaning Assessment. Eight 3-year-olds were excluded[11] for responding only yes or no to all trials in the Meaning Assessment task. Thus, 33 3-year-olds (M = 40.39 months, SD= 3.08 months, female = 15, HA = 16) were included in the analyses for Day 1's Meaning Assessment.

Correct identifications/hits. Planned comparisons for 3-year-olds found result patterns that differed from those of previous studies. In Experiments 1 and 2, 3-year-olds in the HA condition made significantly more hits than those in the LA condition. However, here, the mean number of hits made by the HA and LA 3-year-olds did not differ ($M_{HA} = 3.19$, $SD_{HA} = 1.64$; $M_{LA} = 3.41$, $SD_{LA} = 1.84$; $p = .36$, one-tailed t-test). In fact, there was a non-significant trend for 3-year-olds in the LA condition to make more hits than chance (chance = 2.5 hits), $p = .06$, two-tailed t-test; whereas 3-year-olds in the HA condition performed at chance rates, $p = .12$, two-tailed t-test. However, the number of 3-year-olds to make significantly more hits than chance (correctly identify all five chromium objects) did not differ between the HA group (6 out of the 16 [37.5%] children) and LA group (8 out of the 17 [47.1%] children, $p = .58$, chi-square test. Figure 11 shows the mean number of hits made by 3-year-olds in Experiment 4.

[11] Of the eight excluded children, two of them showed a no bias (HA = 2) and six of them showed a yes bias (HA = 3).

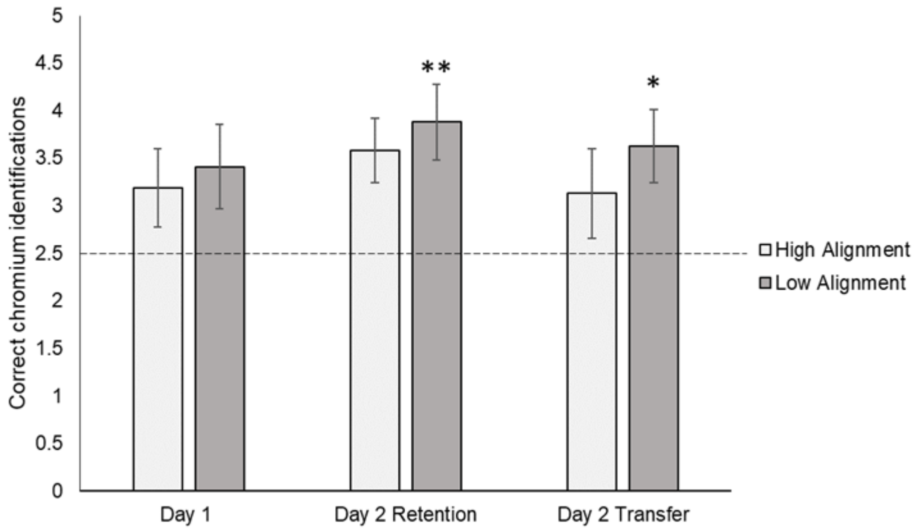


Figure 11. Mean number of chromium objects (max.= 5) correctly endorsed as *chromium* (hits) by 3-year-olds in Experiment 4 Day 1 Meaning Assessment task, Day 2 Retention task, and Day 3 Transfer task. Error bars depict +/- one standard error. The dashed line represents chance performance. * $p < .05$, ** $p < .01$.

Memory check. To better understand the puzzling 3-year-old findings, we compared children's endorsement rates for the two seen chromium objects and the three new chromium objects separately. For the two seen chromium objects, we found a non-significant trend for 3-year-olds in the LA condition to outperform those in the HA condition ($M_{LA} = 1.53$, $SD_{LA} = 0.72$, $M_{HA} = 1.13$, $SD_{HA} = 0.89$; $t(31) = -1.45$, $p = .08$, $d = -0.50$, one-tailed t-test). 3-year-olds in the LA condition correctly endorsed more seen chromium objects than expected by chance (chance = 1 hit), $p < .01$, two-tailed t-test. However, 3-year-olds in the HA condition performed at chance, $p = .58$, two-tailed t-test. Next, we examined children's endorsement of the three new objects. There was no difference between the number of new chromium items endorsed in the two conditions (M_{HA} = 2.06, SD_{HA} = 1.00, M_{LA} = 1.88, SD_{LA} = 1.27; p = .33, one-tailed t-test). Both groups performed at chance rates (chance = 1.5 hits), both ps > .70, two-tailed t-tests.

Incorrect identifications/false alarms. As in previous studies, children did not make more false alarms than expected by chance (chance = 4.5 false alarms). There was a non-significant trend for 3-year-olds in the HA condition (M = 3.19, SD = 2.54) to make fewer false alarms than chance, p = .06, two-tailed t-test. The mean number of false alarms made by 3-year-olds in the LA condition (M = 3.71, SD = 3.14) did not differ from chance, p = .31, two-tailed t-test.

d prime. Combining hits and false alarms to calculate the *d*' scores, we found that 3-year-olds showed low sensitivity to the meaning of *chromium*. *d'* scores were below 1, both for children in the HA condition (M = 0.79, SD = 1.16) and those in the LA condition (M = 0.75, SD = 1.19), with no difference between the two groups, p = .46, one-tailed t-test. Figure 12 shows the mean *d'* scores made by 3-year-olds in the HA and LA conditions.

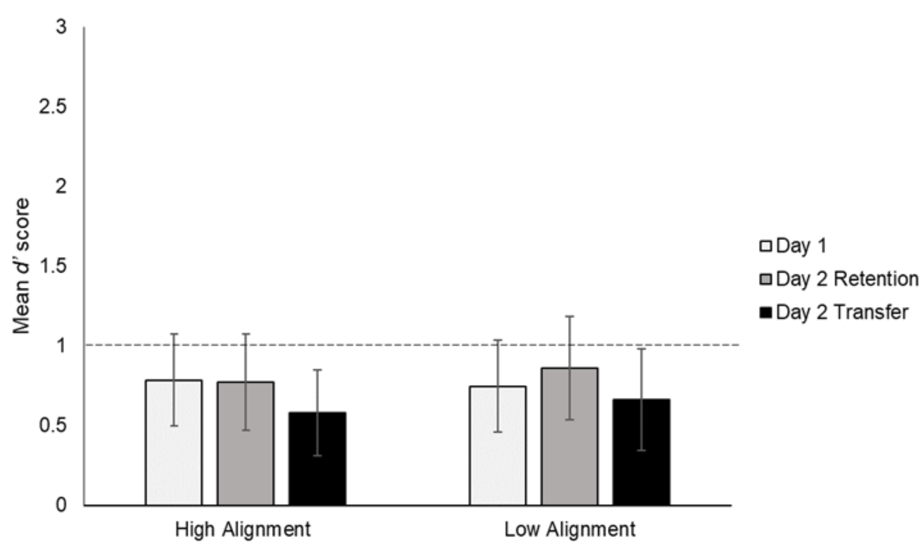


Figure 12. Mean *d'* scores by 3-year-old in Experiment 4. Error bars depict +/2 one standard error. The dashed line represents a d' score of 1.00.

Day 2 Results.

Retention Task. Since we did not find signs of word learning for 3-year-olds in either the HA or LA condition, we did not expect children's performance on Day 2 to be influenced by their performance on Day 1. Thus, we included the children who responded all-yes or all-no on Day 1[12]. As in all previous studies, we excluded the children who answered all-yes or all-no for the assessment of interest here—Day 2's Retention task. Six 3-year-olds (HA = 2) were excluded for this reason. The final analyses for Day 2's Retention task included 35 3-year-olds (M = 40.54 months, SD= 3.21 months, female = 15, HA = 20).

Correct identifications/hits. Planned comparisons between the HA and LA groups found no significant difference (M_{HA} = 3.58, SD_{HA} = 1.47, M_{LA} = 3.88, SD_{LA} = 1.59), p = .29, one-tailed t-test. Children in both conditions made significantly more hits than chance (chance = 2.5 hits), all ps < .01, two-tailed t-tests). We found a similar pattern of results when examining performance on the individual level. There were 8 out of 19 3-year-olds (42.1%) in the HA condition and 9 out of 16 3-year-olds (56.3%) in the LA condition to correctly identify all five chromium objects, with no difference between the two groups, p = .40, chi-square test. Figure 11 shows the mean number of hits made by 3-year-olds in the HA and LA conditions.

Incorrect identifications/false alarms. The number of false alarms made by 3-year-olds in both the HA condition (M = 3.79, SD = 2.76) and LA condition (M = 4.13, SD = 3.07) did not differ from chance (chance = 4.5 false alarms), both ps > .27, two-tailed t-test.

d prime. We found no differences between the mean *d'*s of 3-year-olds in the HA condition (M = 0.80, SD = 1.31) and 3-year-olds in the LA condition (M = 0.86, SD = 1.29), p

[12] This is the same criterion that we employed for the 4-year-olds; however, none of the 4-year-olds answered all-yes or all-no in Day 1's Meaning Assessment task.

= .44, one-tailed *t*-test. In both conditions, children had l *d'* scores below 1, indicating low ability to discriminate between chromium and non-chromium objects. In other words, these children showed low sensitivity to the meaning of the word *chromium*. Figure 12 shows the mean *d'* scores made by 3-year-olds in the HA and LA conditions.

Reminder. When shown two fish and asked to "point to the chromium one, not the blue one," most 3-year-olds (37 out of 41 children[13]) chose the correct chromium object, consistent with previous results on the Initial Exposure Task.

Transfer task. We excluded children who did not point to the chromium object during Day 2 reminder and children who answered yes to all trials in the Transfer task[14]. This resulted in a remaining total of 32 3-year-olds (M = 40.19, SD = 2.93, female = 13, HA = 16).

Correct identifications/hits. Planned comparisons found no differences between the HA and LA conditions (M_{HA} = 3.13, SD_{HA} = 1.89; M_{LA} = 3.63, SD_{LA} = 1.54), p = .21, one-tail *t*-test. 3-year-olds in the LA condition made significantly more hits than chance (chance = 2.5 hits), p = .01, two-tailed *t*-test; but 3-year-olds in the HA condition did not differ from chance, p = .10, two-tailed *t*-test (see Figure 11). There were 6 out of 15 (40.0%) 3-year-olds in the HA condition and 5 out of 16 (31.3%) 3-year-olds in the LA condition to correctly identify all five of the chromium objects. We found no difference between the HA and LA conditions in the number of children to reach the hits criterion, p = .71, chi-square test.

Incorrect identifications/false alarms. The mean number of false alarms made by 3-year-olds in the HA condition (M = 3.00, SD = 2.34) was significantly lower than chance (chance =

[13] Among the four 3-year-olds who failed to point to the chromium object, HA= 3.

[14] Five children were excluded for answering yes to all trials in the Transfer task (HA = 2). An additional three children (HA =2) answered no to all of the trials in the Transfer task but was included in the final analyses, since answering all no was an indicator of failing to transfer.

4.5 false alarms), $p = .02$, two-tailed *t*-test; but the number of false alarms made by 3-year-olds in the LA condition ($M = 4.69$, $SD = 3.50$) did not differ from chance, $p = .83$, two-tailed *t*-test.

d prime. As in Day 1's Meaning Assessment and Day 2's Retention task, 3-year-olds in both the HA and LA conditions had *d'* scores below 1, indicating low discrimination of chromium vs non-chromium objects. There were no differences between the mean *d'*s of 3-year-olds in the HA condition ($M = 0.85$, $SD = 1.11$) and those in the LA condition ($M = 0.53$, $SD = 1.31$), $p = .24$, one-tailed *t*-test. Figure 12 shows the mean *d'* scores made by 3-year-olds in the HA and LA conditions.

Comparison of all tasks. We conducted a repeated-measures ANOVA with the 26 children who were included in the analyses for all three tasks[15]. We used Hits as the dependent variable, and Task (Day 1 Meaning Assessment, Day 2 Retention, and Day 2 Transfer) as a within-subject variable and Alignment (HA and LA) as a between-subject variable. There were no significant effects, all *p*s > . 30.

Discussion of 3-year-olds' performance.

The findings for 3-year-olds in Experiment 4 were puzzling. Based on Experiment 1 and 2's findings, we had predicted that 3-year-olds in the HA condition in Experiment 4 would perform well on Day 1's Meaning Assessment task. However, their performance neither differed from chance nor the performance of 3-year-olds in the LA condition. Since we did not find a difference between 3-year-olds in the HA and LA conditions on Day 1, we did not expect to find any differences on the following tasks. Indeed, 3-years-olds in the HA and LA conditions performed comparably on both of Day 2's tasks.

[15] We also found no differences between Task or Alignment when we included all 3-year-olds (N = 41).

Why did 3-year-olds in the HA condition fail to link the novel word and color on Day 1? We speculate that the unfamiliar shapes (crosses and upside-down triangles) of the Initial Exposure objects may have impeded 3-year-olds' initial mapping. More cognitive resources may have been directed towards processing the objects themselves, and this focus on individual objects may have lessened the likelihood of simultaneously processing both objects, and therefore the likelihood of comparison. Even if spontaneous comparison did occur for some children, such that the novel color was highlighted, it could be that the encoding demands reduced cognitive resources, lessening the likelihood of forming a link between the word and the novel color.

Across the Meaning Assessment task, Retention task, and Transfer task, 3-year-olds in both the HA and LA conditions had low *d'* scores—in fact, their scores were consistently below 1.00. Table 1 shows the *d' s*cores for children in both the HA and LA conditions in Experiment 4. Three-year-olds had difficulties discriminating between the *chromium* and non-*chromium* objects in all three tasks, suggesting that throughout the experiment, they failed to link the novel color with the new word. Thus, although some 3-year-olds made significantly more hits than chance on certain tasks, it is likely due to reasons other than learning.

Table 1.
Mean *d'* score of 4-year-olds and 3-year-olds in the High Alignment (HA) and Low Alignment (LA) conditions in Experiment 4. Standard deviations in parentheses.

	4-year-olds		3-year-olds	
	HA	LA	HA	LA
Day 1 Meaning Assessment	2.20 (0.86)	1.57 (1.46)	0.79 (1.16)	0.75 (1.19)
Day 2 Retention Task	2.19 (0.84)	1.79 (1.43)	0.80 (1.31)	0.86 (1.29)
Day 2 Transfer Task	2.22 (1.04)	1.61 (1.42)	0.85 (1.11)	0.53 (1.31)

We speculate that 3-year-olds and 4-year-olds showed high hit rates on Day 2's Retention task for different reasons. Since 3-year-olds did not learn the meaning of *chromium* on Day 1, they may have been confused and thus showed an overall tendency to respond positively in the Retention task. In addition to high hit rates, 3-year-olds also had relatively high false alarm rates (M_{HA} = 3.79, M_{LA} = 4.13), especially in comparison to the 4-year-olds (M_{HA} = 1.42, M_{LA} = 1.55). This resulted in low *d'* scores, as shown in Table 1. In contrast, 4-year-olds' high performance on Day 2's Retention task may be due to a test-retest effect. Although none of the children were given any corrective feedback, the Retention task—an exact replica of Day 1's Meaning Assessment task—may have elicited retrieval of Day 1's events. In addition, additional exposure to the color and word may have also promoted the 4-year-old's performance.

Unlike the Retention task, high performance on the Transfer task required children to have some understanding of the mapping between the word *chromium* and the color. Three-year-olds, who did not seem to make the mapping on Day 1, also performed poorly on Day 2's Transfer task, as evidenced by their low *d'* scores. Four-year-olds in both the HA and LA condition performed well on Day 1's Meaning Assessment task; however, there was still an advantage of high-alignment over low-alignment. This advantage was amplified in the Transfer task—only 4-year-olds in the HA condition was able to correctly identify new chromium stimuli (fish) as significantly above chance rates. This suggests that learning through indirect situations may have been of higher quality for children who formed a stronger initial mapping (4-year-olds who saw easy-to-align perceptual cues) than those who formed a weaker initial mapping (4-year-olds who saw hard-to-align perceptual cues).

Chapter 5: Discussion of the Perceptual Alignment Studies

Experiments 1-4 tested the hypothesis that the presence of optimal perceptual comparisons contributes to referential transparency in incidental learning. The results support the claim that spontaneous comparison processes promote incidental word learning. Experiment 1 established the basic logic of the studies. Children were given either highly alignable objects—two squares, one chromium and one blue, or hard-to-align objects—a chromium circle and a blue square. We asked the children to "point to the chromium one, not the blue one" (as in Carey and Bartlett's (1978) study). Although 96% of the children pointed correctly across studies (and those who did not do so were omitted from further analyses), the children differed sharply in how much they learned from this experience. When given a yes-no "Is this one chromium?" task (the Meaning Assessment task) ten minutes later, the high-alignment group was significantly better at identifying chromium objects than was the low-alignment group, across both ages.

In Experiment 2, we addressed a potential concern. Perhaps children were engaged in hypothesis-testing during the pointing task, with the goal of trying to understand the term *chromium*. In this case, the low-alignment group was at a disadvantage. Whereas the high-alignment group saw two figures that differed only in color, the low-alignment group saw two figures that differed in both shape and color. Thus, they could have formed two hypotheses—a shape hypothesis ("chromium means square") and a color hypothesis ("chromium means dark green"). In contrast, the high-alignment group would have only one possible hypothesis—the color hypothesis—and would naturally do better at deriving the meaning of chromium. To test this possibility, in Experiment 2, we presented 3-year-olds with two initial exposures, designed so that both groups would have adequate information to rule out a shape hypothesis and arrive at the correct color hypothesis (see Figure 1). However, we once again found that children in the high-alignment condition outperformed those in the low-alignment condition in the Meaning Assessment. This argues against the possibility that that children were engaged in deliberate hypothesis-testing—rather, spontaneous comparison processes made the color contrast obvious in the high-alignment condition.

In Experiment 3, we contrasted the implicit learning conditions of Experiments 1 and 2 with direct, explicit word-learning instructions. Children were told at the start of the study that they should try to learn the word *chromium*. The rest of the study was as in Experiment 1: children received a brief Initial Exposure pointing task, followed later by a yes-no Meaning Assessment task. Given direct instructions, children learned equally well across the high-alignment and low-alignment conditions. Cross-study comparisons between direct (Experiment 3) and indirect (Experiment 1) confirmed that the low-alignment group performed significantly better in the direct than in the indirect condition. For the high-alignment group, there was no difference between the two conditions.

These findings reinforce our claim that children in our indirect studies (Experiments 1, 2, and 4) were not engaged in deliberate word learning. When directly asked to learn the word, the low-alignment group performed as well as the high-alignment group. This means that the low-alignment condition is adequate for word learning, *if* children see the task as word-learning. Since children in the low-alignment condition were quite poor at learning the word under indirect conditions, we infer that children in these studies were not engaged in trying to learn the word. This in turn supports our claim that the high-alignment advantage in the indirect studies stemmed from spontaneous, non-intentional comparison processes.

Experiment 4 showed that the indirectly learned word (*chromium*) was retained over a delay and could be transferred to new kinds of objects. Four-year-olds in both conditions performed above chance on retention and transfer after a two-day delay. However, the advantage of high-alignability persisted—4-year-olds given highly alignable pairs were better able to transfer the meaning of *chromium* to a new set of objects then those given less alignable pairs.

Perceptual comparison contributes to referential transparency

Recent research has emphasized that the quality of the input, rather than the sheer quantity, is a strong predictor of children's vocabulary growth (e.g. Cartmill et al., 2013; Hirsh-Pasek et al., 2015; Rowe, 2012). Our findings on the role of perceptual alignability accord with this claim. Seeing one highly alignable pair provided high-quality information that was sufficient for children to learn the word *chromium* from indirect exposure in Experiment 1. In Experiments 2, and 4, children received two opportunities to learn the word *chromium* from indirect input; but even with two opportunities, children in the LA condition were still disadvantaged relative to those in the HA group. This is despite receiving sufficient information across two trials to settle on color as the relevant dimension.

At the same time, the findings of Experiment 3, in which children were given direct instructions to learn the word *chromium*, underline the role of pedagogical factors. When given direct instructions to learn a new word, children in the low alignment condition learned the new word as well as those in the high alignment condition. This brings us to the next investigation: whether and how the language children hear during indirect presentation contributes to learning.

Chapter 6: Beyond Perceptual Alignment—the Role of Linguistic Cues

In Experiments 1-4, we manipulated the alignability of the objects presented to the participants while keeping the linguistic cues consistent. All children were asked, "Can you point to the chromium one? The chromium one, not the blue one." Across experiments, we found that when presented with high-quality linguistic cues such as these, children who saw easy-to-align perceptual cues (HA condition) were better at linking a novel word to its referent property than those who saw hard-to-align perceptual cues (LA condition).

Previous research has suggested that the regularities of language also allow learners to make inferences about words independent of perceptual cues. For example, many studies on syntactic bootstrapping have suggested that children are sensitive to the syntactic framing used to present novel words and can use the framing to make inferences about the meaning of the words (e.g. Gleitman, 1990; Fisher, 1996; Mintz & Gleitman, 2002; Wellword, Gagliardi, & Lidz, 2016).

Our previous linguistic cues may have facilitated learning through three different ways: referencing the target object, parallel syntax, and semantic contrast.

First, children were highly accurate at pointing to the chromium object during the Initial Exposure. We propose that this is due to the linguistic cues' clear *reference to the target object*. To comply with the experimenter's request, children merely had to point to the non-blue object ("not the blue one")—without even having to process the meaning of *chromium*. Although directing the children's attention to an object that displays the target property does not solve the mapping problem (there are still more than one potential referent properties; see Chapter 2 Experiment 2), it should greatly increase the likelihood of children making a correct word-to-property mapping.

Second, the linguistic cues included *parallel syntax* between the phrases "the chromium one" and "the blue one". Not only do the two phrases share the same structure of "determiner + adjective + pronoun", both the determiner and pronoun are the same words ("the" and "one"). This high level of repetition may have invited spontaneous comparison between the two utterances and highlighted the alignable difference between the words *chromium* and *blue*. This also sets the stage for the third cue—*semantic contrast*.

The phrase "the chromium one, not the blue one" contains a word-to-word mapping between the novel word *chromium* and a known color word *blue*. Sandhofer and Smith (1999) have argued that learning word-word mappings—e.g. knowing that a specific group of words all refer to colors without actually knowing *which* specific colors—promotes learning word-property mappings. They found that children typically understood that one should answer from a set of specific words (*red, blue*, *orange*, etc.) when asked "what color is this" before they could map these words to specific colors. Furthermore, the age of acquisition of word-word mappings predicted the age of acquisition of word-property mappings. Associating different color words with each other may highlight the fact that there is an unknown dimension to be learned. This may ultimately draw children's attention to the dimension of interest (Sandhofer & Smith, 1999). In our experiments, hearing *chromium* in conjunction with *blue* may have served as a contextual cue to direct children's attention to color.

What are the respective roles of reference to the target object, parallel syntax and semantic contrast in children's indirect property word learning? Previous research has examined the role of parallel syntax (usually in conjunction with highly alignable perceptual cues and linguistic reference to the carrier object). Heibeck and Markman (1987) presented preschoolers with object pairs that had the target property as an alignable difference (analogous to what we show children in the HA condition) and gave them either specific linguistic contrast (e.g. "bring me the chartreuse one, not the red one") or vague linguistic contrast (e.g. "bring me the chartreuse one, not the other one"). If children were making inferences about the meaning of *chartreuse* based on semantic contrast, one would expect that those in the specific contrast condition would show better learning than those in the vague contrast condition. However, Heibeck and Markman (1987) found that children learned equally well in both conditions; thus, they found no evidence for additive benefit of semantic contrast when presented in conjunction with high quality perceptual cues[16].

One possible interpretation for these findings is that the target dimension of color may have already been perceptually obvious in Heibeck and Markman's studies, and that the linguistic cues were redundant. Under this assumption, children would have learned equally well even if they had been given instructions that lacked all three aspects of linguistic support discussed above. Another interpretation is that linguistic cues do benefit children's indirect word learning, but that this effect is mainly driven by inviting comparison through parallel syntax. Parallel syntax ("the chromium one, *not* the other one"). may have increased the likelihood of spontaneous comparison between the two objects. Alternately, even if the dimension of color was already highlighted through perceptual alignment, parallel syntax may have led to deeper processing of the actual color. Under this assumption, children would not learn as well (and may even fail at the initial pointing task) if they were only given semantic cues, without parallel syntax. In the next chapter, we present two experiments examining the unique contributions of semantic contrast on children's indirect word learning.

Chapter 7: Semantic Contrast Facilitates Indirect Property-Word Learning

In this chapter, we present two experiments that explicitly test the role of linguistic cues in children's indirect property word learning. In Experiment 5, we presented children with

---

[16] However. the Heibeck and Markman study had only 16 children, ranging from 2 to 4 years, so it may have been underpowered.

linguistic cues that included both reference to the target object and semantic contrast, but not parallel syntax. In Experiment 6, we further decreased the quality of information in the linguistic cues by retaining only a reference to the target object. By comparing children's learning in Experiment 5 with that in Experiment 1 (in which the linguistic cues included references to the target object, parallel syntax, and semantic contrast) and Experiment 6 (in which the linguistic cues only included reference to the target object), we can explore the unique role of semantic contrast.

In Chapter 4's Experiment 4, we found that 3-year-olds, even those presented with high-alignment perceptual cues, failed to learn the novel word when the Initial Exposure pairs were unfamiliar shapes. We suspect the novel object shapes diverted children's attention, leaving them with fewer cognitive resources. Thus, in the experiments presented below, we returned to the familiar object shapes that we had previously used in Experiments 1-3. However, the nature of our research questions inherently required children to devote more attention to processing the less optimal linguistic cues. To be more prudent, we examined and reported the findings of 3-year-olds and 4-year-olds separately. In this chapter, we present two experiments that explicitly test the role of linguistic cues in children's indirect word learning.

Experiment 5

In Experiment 5 we told children, "I don't like the blue one. Can you point to the one that's chromium?" This linguistic cue retained some degree of transparency regarding the target object (not the blue one, since the experimenter dislikes it). It also retained the semantic contrast between the novel word *chromium* and known color word *blue*. However, it did not include the parallel syntactic structure ("the X one, not the Y one") present in the previous studies. Table 2 shows the linguistic input provided to children in our Experiments 1, 5, and 6, as well as in

Heibeck and Markman (1987). To examine the interaction between semantic contrast and perceptual alignment, we manipulated the alignability of the object pairs that children saw.

Table 2.
Linguistic input presented in Experiments 1, 5, and 6, and Heibeck and Markman (1987).

	Linguistic Input	Included Cues		
		Reference to Target Object	Parallel Syntax	Semantic Contrast
Experiment 1	*Can you point to the chromium one? The chromium one, not the blue one.*	Yes	Yes	Yes
Experiment 5	*I don't like the blue one. Can you point to the one that's chromium?*	Yes	No	Yes
Experiment 6	*Can you point to the chromium one? The chromium one, not that one.* + Wave of hand	Yes	No	No
Heibeck & Markman (1987) Specific Contrast	*Bring me the chartreuse one, not the red one.*	No	Yes	Yes
Heibeck & Markman (1987) Vague Contrast	*Bring me the chartreuse one, not the other one.*	No	Yes	No

Methods.

Participants. A total of 40 3-year-olds (*M* = 41.88 months, *SD* = 2.78 months, female = 22) and 34 4-year-olds (*M* = 54.44 months, *SD* = 2.56 months, female = 18) participated in this study. Children were randomly assigned to either the High Alignment (HA) or Low Alignment (LA) condition. Ten additional children were tested but excluded, three for failing to complete the study, one for parental interference, one for taking an unscheduled break during the study, one for failing to reach criterion for the yes-no sorting task[17], and four for answering no to all trials in the Meaning Assessment task[18]. All children were recruited through the same methods as previous studies and received a small gift at the end of the experiment.

[17] The yes-no classification task was used as a warm-up for the Meaning Assessment task. To be included in the final analyses, children need to be correct on at least eight of the twelve yes-no classification trials. See Chapter 2 for more details.
[18] The four children who answered no to all trials in the Meaning Assessment included one 3-year-old in the HA condition, one 3-year-old in the LA condition, and two 4-year-olds in the LA condition.

Materials and procedure. The procedure and materials for Experiment 5 were similar to those of Experiment 1, except for the linguistic cues presented to children during the Initial Exposure. In Experiment 5, we asked children, “I don’t like the blue one. Can you point to the one that’s chromium?” The new linguistic cues preserved the semantic contrast between the novel word (*chromium)* and a known color word (*blue*), while removing the phrasal contrast of “the X one, not the Y one”. As before, children could comply to the experimenter’s request without understanding the meaning of *chromium* by simply pointing to the non-blue object.

Otherwise, the procedure and materials were the same as those of Experiment 1. Children were presented with only one exposure to the novel word and color, with the exposure embedded in a pointing task. After a short break of approximately ten minutes, participants were assessed on how well they learned the new word based on their performance on a yes-no sorting task—the Meaning Assessment. The Meaning Assessment task, our key measure to whether children learned the word or not, consisted of the same objects as in Experiment 1. See Figure 13 for schematics of the stimuli.

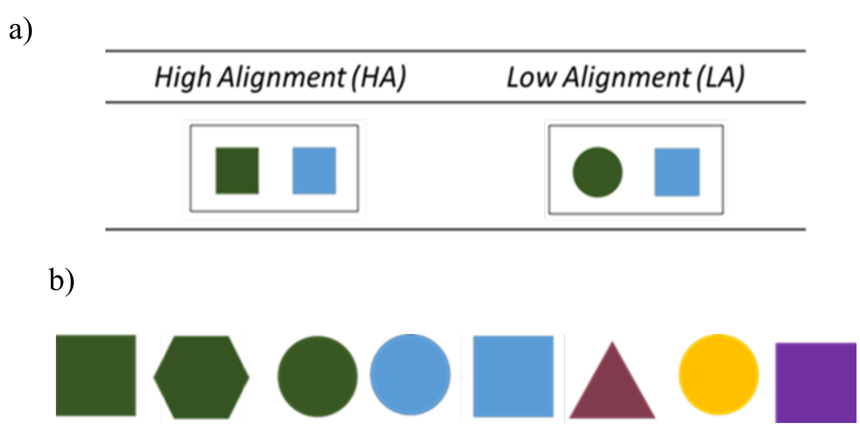


Figure 13. Schematics of objects presented in Experiments 5 and 6 in (a) the Initial Exposure; and (b) the Meaning Assessment task.

Results for 4-year-olds.

Initial Exposure. As in previous studies, 4-year-olds performed well on the Initial Exposure. Among the 34 children who participated in Experiment 5, 33 of them (97.1%) correctly pointed to the chromium object despite being presented with less optimal linguistic cues. The child who failed to do so was excluded from further analyses.

Meaning Assessment. We scored the number of hits (chromium objects correctly endorsed; max. = 3) and false alarms (non-chromium objects incorrectly endorsed, max. = 5) made by each child. Figure 14 shows the mean number of hits and false alarms made by 4-year-olds in Experiment 5.

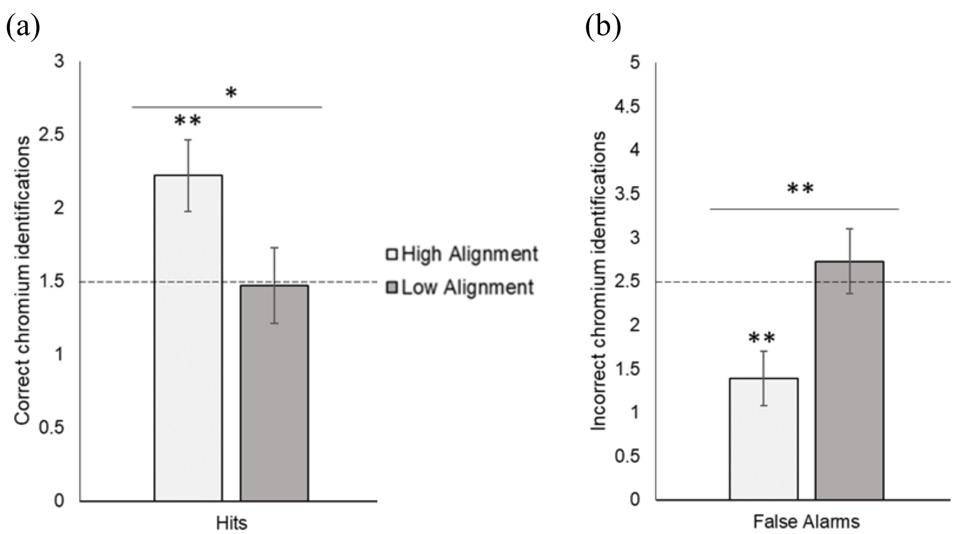


Figure 14. (a) Mean number of chromium objects (max.= 3) correctly endorsed as *chromium* (hits); and (b) mean number of non-chromium objects (max. =5) incorrectly endorsed as *chromium* (false alarms) by 4-year-olds in Experiment 5's Meaning Assessment task . Error bars depict +/- one standard error. The dashed lines represent chance performance. * $p < .05$, ** $p < .01$.

Correct identifications/hits. Planned comparisons found that 4-year-olds in the HA condition made significantly more hits than those in the LA condition ($M_{HA} = 2.22$, $SD_{HA} = 1.00$; $M_{LA} = 1.47$, $SD_{LA} = 0.99$), $t(31) = 2.17$, $p = .02$, $d = 0.75$, one-tailed t-test. In addition, only the

HA group made significantly more hits than expected by chance (chance = 1.5 hits), $t(17) = 3.05$, $p = .007$, $d = 0.72$; two-tailed *t*-test; the number of hits made by the LA group did not differ from chance, $t(14) = -0.13$, $p = .90$, $d = -0.03$, two-tailed *t*-test.

We found similar patterns of results when examining performance at the individual level. As in Experiment 1, we set the hit criterion at two or more correct (out of three). We found that 15 out of 18 children (83.3%) in the HA condition and 6 out of 15 children (40.0%) in the LA condition reached the hit criterion. The number of 4-year-olds who reached criterion in the HA condition was significantly higher than that of the LA condition, $X^2(1, N = 33) = 6.64$, $p = .01$, $\varphi = 0.45$, chi-square test.

Memory check. We next examined children's endorsement of the previously seen Initial Exposure object and the two new chromium objects. We found similar patterns of results as with Experiment 1.The number of children to correctly endorse the seen object did not differ between the HA condition (15 out of 18 [83.3%]) and LA condition (11 out of 15 [73.3%]), $p = .48$, chi-square test. However, children in the HA condition were more likely to endorse new chromium objects ($M = 1.39$, $SD = 0.78$) than those in the LA condition ($M = .73$, $SD = 0.80$), $t(31) = 2.38$, $p = .01$, $d = 0.81$, one-tailed *t*-test. In addition, 4-year-olds in the HA condition correctly endorsed significantly more seen objects than expected by chance (chance = 1; $t(17) = 2.12$, $p = .05$, $d = 0.50$; two-tailed *t*-tests), but 4-year-olds in the LA condition did not ($p = .22$, two-tailed *t*-test).

Incorrect identifications/false Alarms. To examine whether the difference in hits was due to an overall yes bias, we compared the number of false alarms made by each group to chance (chance = 2.5 false alarms). Four-year-olds in the HA condition made significantly fewer false alarms than chance ($M = 1.39$, $SD = 1.29$), $t(17) = -3.66$, $p = .002$, $d = -0.86$; two-tailed *t*-test;

while 4-year-olds in the LA group (*M* = 2.73, *SD* = 1.28) did not differ from chance, *p* = .49, two-tailed *t*-test.

Results for 3-year-olds.

Initial Exposure. Among the 40 3-year-olds who participated in Experiment 5, 34 of them (85.0%) correctly pointed to the *chromium* object[19]. Three children in the HA condition and three children in the LA condition failed to do so. As before, children who did not correctly point to the chromium object during the Initial Exposure were excluded from further analysis.

Meaning Assessment. As before, we scored the number of hits (chromium objects correctly endorsed; max. = 3) and false alarms (non-chromium object incorrectly endorsed; max. = 5) made by each child. Figure 15 shows the mean number of hits and false alarms made by 3-year-olds in Experiment 5.

Correct identifications/hits. No difference was found between the mean number of hits made by 3-year-olds in the HA condition (*M* = 2.18, *SD* = 1.02) and those in the LA condition (*M* = 2.06, *SD* = 1.09), *t*(32) = 0.22, *p* = .37, *d* = 0.11, one-tailed *t*-test. Three-year-olds in both the HA and LA conditions made significantly more hits than expected by chance (chance = 1.5 hits; HA: *t*(16) = 2.75, *p* = .01, *d* = 0.67; LA: *t*(16) = 2.12, *p* = .05, *d* = 0.51, one-tailed *t*-tests). There was also no difference in the number of 3-year-olds to reach the hit criterion in the HA condition (12 out of 17 [70.6%]) and the LA condition (12 out of 17), *p* = 1.00, chi-square test.

[19] The proportion of 3-year-olds to pass the Initial Exposure in Experiment 5 (34 out of 40 children) was not significantly different from that of Experiment 1 (61 out of 65 children), $X^2(1, N = 105) = 2.25$, *p* = .13, $\varphi = .15$.

(a) (b)

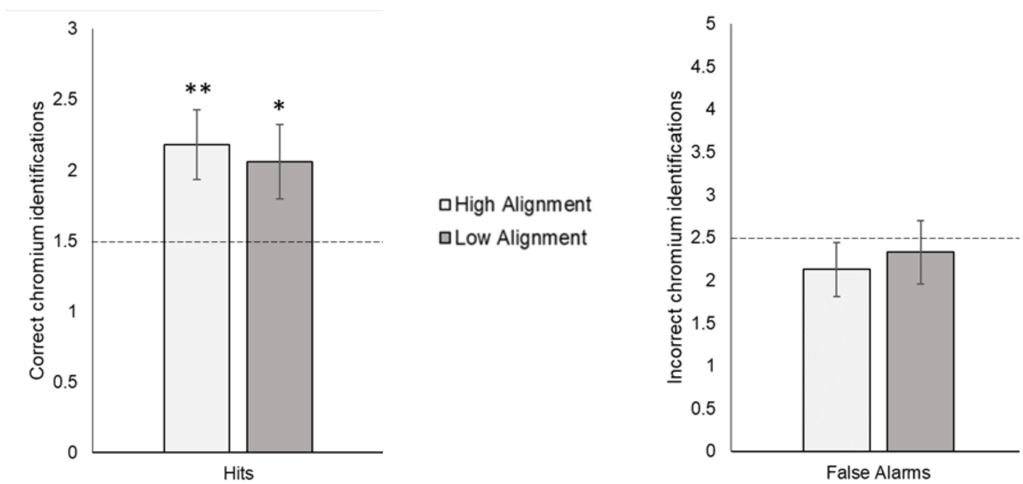


Figure 15. (a) Mean number of chromium objects (max.= 3) correctly endorsed as *chromium* (hits); and (b) mean number of non-chromium objects (max. =5) incorrectly endorsed as *chromium* (false alarms) by 3-year-olds in Experiment 5's Meaning Assessment task . Error bars depict +/- two standard errors. The dashed lines represent chance performance. * $p < .05$, ** $p < .01$.

Memory check. The number of children to correctly endorse the seen object did not differ between 3-year-olds in the HA condition (11 out of 17) and LA condition (12 out of 17 [70.6%]), $p = .71$, chi-square test. The mean number of new chromium objects (max. = 2) endorsed by 3-year-olds in the HA condition ($M = 1.53$, $SD = .72$) and LA condition ($M = 1.35$, $SD = .86$) also did not differ ($t(32) = .65$, $p = .26$, $d = 0.23$, one-tailed *t*-test). However, the mean number of new chromium objects correctly endorsed by 3-year-olds in the HA condition was significantly higher than chance ($t(16) = 3.04$, $p = .008$, $d = 0.74$; two-tailed *t*-tests), whereas it was not the case for 3-year-olds in the LA condition ($p = .11$, two-tailed *t*-test).

Incorrect identifications/false alarms. Neither the mean number of false alarms made by 3-year-olds in the HA condition ($M = 2.24$, $SD = 1.30$) and that made by those in the LA condition ($M = 2.35$, $SD = 1.62$) differed from chance (both *p*s > .40, two-tailed *t*-tests). Although a *d'* analysis is not possible with the low number of items, the relatively high number of false alarms among 3-year-olds raises a concern that the high hit rate may reflect a greater propensity to say “yes” rather than greater ability to discriminate the chromium color.

Experiment 5 Discussion.

In Experiment 5, we examined the role of semantic contrast in children’s indirect property word learning. We presented 3-and 4-year-olds with the linguistic cue “I don’t like the blue one. Can you point to the one that’s chromium?”, while manipulating the perceptual alignment of the object pairs that children saw. Compared to the linguistic cues presented the previous studies (“Can you point to the chromium one? The chromium one, not the blue one.”), the new instruction retained the semantic contrast between the novel word *chromium* and a known color *blue.* However, it lacked parallel syntax between “the chromium one” and “not the blue one”, and this might have diminished the likelihood of children spontaneously noting the contrast between the two phrases.

Interestingly, 3-year-olds and 4-year-olds responded differently to the new linguistic cues. The performance of 4-year-olds largely paralleled that of Experiment 1. We found that 4-year-olds in the High Alignment demonstrated better learning of the novel property word (both high hit rates and low false-alarm rates) than 4-year-olds in the Low Alignment condition. However, 3-year-olds demonstrated equally high hit rates in the High Alignment and Low Alignment conditions—a stark contrast to the findings of Experiment 1. We will return to this point in this chapter’s overall discussion.

Experiment 6

In Experiment 6, we further decreased the quality of information in the linguistic cues. We asked children, “Can you point to the chromium one—the chromium one, not *that* one” while dismissing the non-target object with a wave of the hand. This new linguistic cue did not include a known color word; thus, children would not have been able to use semantic contrast to infer the meaning of the novel word *chromium*. The new linguistic cues also lacked the parallel syntactic structure of "the (blank) one". To retain the reference to the carrier object we added the gestural cue. The idea was to give children enough information to know which object to choose (the one not dismissed) without their having to understand the meaning of *chromium*.

Methods.

Participants. A total of 41 4-year-olds (M = 53.39 months, SD = 3.84 months, females= 19) and 43 3-year-olds (M = 42.28 months, SD = 2.68 months, females= 24) participated in this study. An additional 17 children were tested but excluded from further analyses, one due to experimenter error, one child for taking an unscheduled break during the task, two children for failing to complete the task, two children for failing to reach the criterion for the yes-no sorting task, and eleven children for answering all yes or all no in the Meaning Assessment task[20] As in previous studies, children were randomly assigned to see object pairs that were either high in alignment (HA condition) or low in alignment (LA condition).

Materials and procedure. We modified the linguistic cues that children received during the Initial Exposure to eliminate the presence of a known color word. On the crucial chromium exposure trial, children were asked "Can you point to the chromium one? The chromium one, not *that* one." Upon saying "not that one", the experimenter dismissively waved across the incorrect object (a blue circle or blue square) so that children could still infer which of the two objects was the target. Otherwise, the procedure and materials for Experiment 6 were the same as those for Experiment 1 (and Experiment 5).

[20] Among these eleven children were eight 3-year-olds and three 4-year-olds.

Results for 4-year-olds.

Initial Exposure. Of the 41 4-year-olds who participated in Experiment 6, 37 of them (90.2%) correctly pointed to the chromium object during the Initial Exposure. Four children (HA = 2) failed to do so and were excluded from further analyses.

Meaning Assessment. As in previous studies, we examined the mean number of hits (correctly endorsed chromium objects; max. = 3) and mean number of false alarms (incorrectly endorsed non-chromium objects; max. = 5) made by children in the HA and LA conditions. Figure 16 shows the mean number of hits and false alarms made by 4-year-olds in Experiment 6.

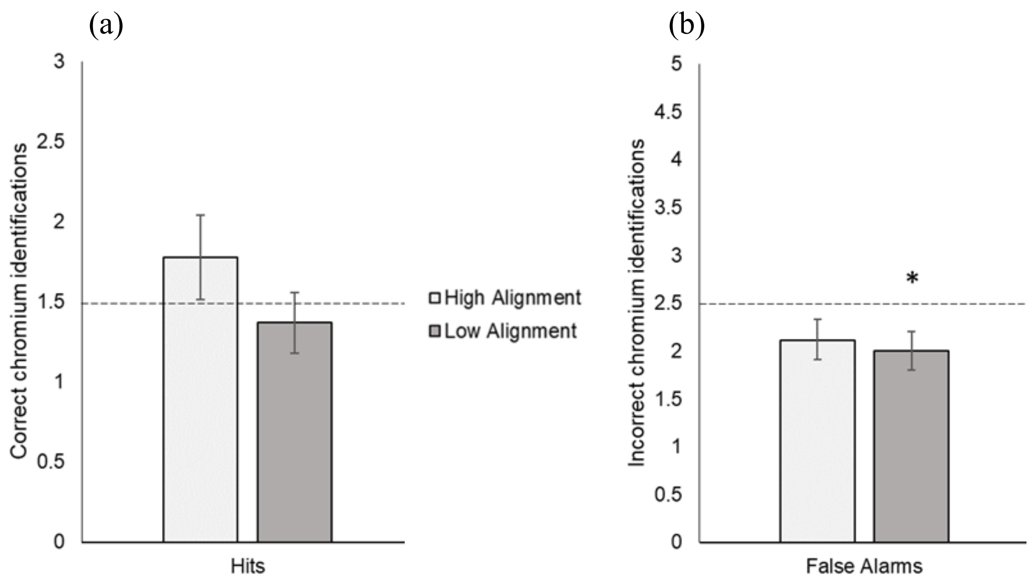


Figure 16. (a) Mean number of chromium objects (max.= 3) correctly endorsed as *chromium* (hits); and (b) mean number of non-chromium objects (max. =5) incorrectly endorsed as *chromium* (false alarms) by 4-year-olds in Experiment 6's Meaning Assessment task . Error bars depict +/- one standard error. The dashed lines represent chance performance. * $p < .05$.

Correct identifications/hits. Planned comparisons found no difference between the mean number of hits made by 4-year-olds in the HA condition ($M = 1.78$, $SD = 1.11$) and LA condition ($M = 1.37$, $SD = 0.83$), $p = .11$, $d = 0.42$, one-tailed t-test. Four-year-olds in both conditions performed at chance rates (chance = 1.5 hits), both ps > .31, two-tailed t-tests. There was also no difference between the number of children who reached the hit criterion (two or more out of three) in the HA condition (11 out of 18 [61.1%]) and the LA condition (7 out of 19 [36.8%]), $p = .14$, chi-square test.

Memory check. We also examined children's endorsement of the previously seen Initial Exposure object and two new chromium objects. The number of children to correctly endorse the seen object did not differ between 4-year-olds in the HA condition (13 out of 18 [72.2%]) and those in the LA condition (14 out of 19 [73.7%]), $p = .92$, chi-square test. However, there was a non-significant trend for children in the HA condition to endorse more new chromium objects ($M = 1.06$, $SD = 0.94$) than their age-matched peers in the LA condition ($M = 0.63$, $SD = 0.76$), $t(35) = 1.51$, $p = .07$, $d = 0.25$, one-tailed t-test.

Incorrect identifications/false alarms. Four-year-olds in the LA condition ($M = 2.00$, $SD = 0.88$) made significantly fewer false alarms than expected by chance (chance = 2.5 false alarms), $t(18) = -2.47$, $p = .02$, $d = -0.57$, two-tailed t-test, but 4-year-olds in the HA condition ($M = 2.11$, $SD = 0.63$) did not differ from chance, $p = .11$, two-tailed t-test.

Results for 3-year-olds.

Initial Exposure. Of the 43 3-year-olds who participated in Experiment 6, only 32 of them (74.4%) correctly pointed to the chromium object during the Initial Exposure; eleven children (HA = 4) failed to do so. The proportion of children who were correct was still significantly higher than expected if children were randomly selecting ($p < .001$, binomial test) between the Initial Exposure objects. However, this proportion was significantly lower than that of Experiment 1 ($X^2(1, N = 108) = 8.17$, $p = .004$, $\varphi = .28$, chi-square test), although it did not differ from that of Experiment 5 ($X^2(1, N = 83) = 1.43$, $p = .23$, $\varphi = .13$, chi-square test). This is not due to a failure to understand the pointing task—children's accuracy rates on all three filler trials were around 95%. We excluded children who did not point to the chromium object during the Initial Exposure from further analysis.

Meaning Assessment. We examined the number of correct identifications (hits) and incorrect identifications (false alarms) made by children in the HA and LA conditions. Figure 17 shows the mean number of hits and false alarms made by 3-year-olds in Experiment 6.

Correct identifications/hits. Planned comparisons found no differences between the mean number of hits made by 3-year-olds in the HA condition and LA condition, $p = .21$, one-tailed t-tests. However, there was a non-significant trend for 3-year-olds in the HA condition to make more hits than chance ($M = 2.00$, $SD = 1.09$,) $t(17) = 1.96$, $p = .07$, $d = 0.46$, two-tailed t-test; while 3-year-olds in the LA condition did not differ from chance ($M = 1.71$, $SD = 0.83$), $t(13) = .97$, $p = .35$, $d = 0.26$, two-tailed t-test. The number of children to reach the hit criterion (two or more out of three) also did not differ between 3-year-olds in the HA condition (12 out of 18 [66.7%]) and LA condition (7 out of 14 [50.0%]), $p = .34$, chi-square test.

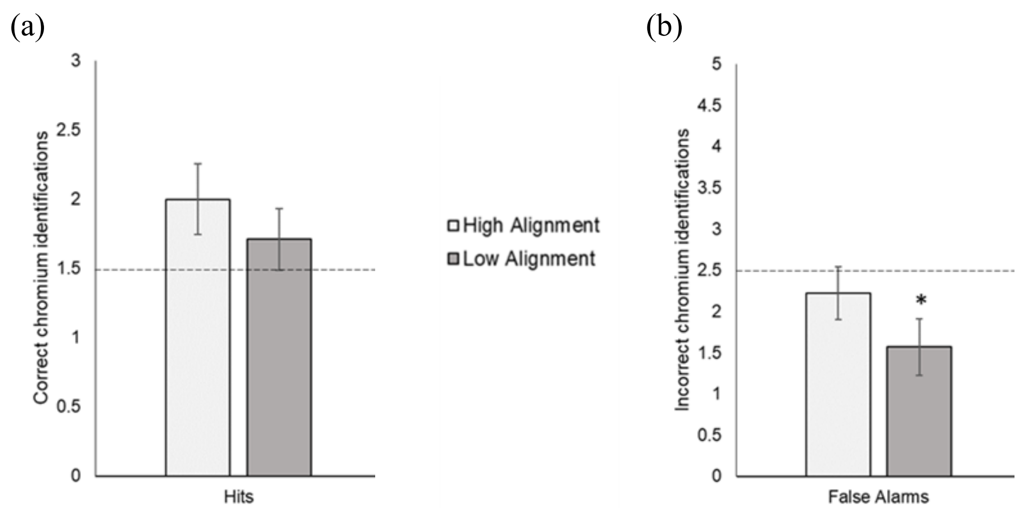


Figure 17. (a) Mean number of chromium objects (max.= 3) correctly endorsed as *chromium* (hits); and (b) mean number of non-chromium objects (max. =5) incorrectly endorsed as *chromium* (false alarms) by 3-year-olds in Experiment 6's Meaning Assessment task . Error bars depict +/- one standard error. The dashed lines represent chance performance. * $p < .05$.

Memory check. We examined children's endorsement of the previously seen Initial Exposure object and two new chromium objects. The number of children to correctly endorse the seen object did not differ between 3-year-olds in the HA condition (12 out of 18 [66.7%]) and LA condition (11 out of 14 [78.6%]), $p = .46$, chi-square test. However, there was a non-significant trend for children in the HA condition to endorse more new chromium objects ($M = 1.33$, $SD = 0.84$) than those in the LA condition ($M = 0.93$, $SD = 0.83$), $t(30) = 1.36$, $p = .09$, $d = 0.24$, one-tailed t-test.

Incorrect identifications/false alarms. Three-year-olds in the LA condition made significantly fewer false alarms than expected by chance $t(13) = -2.71$, $p = .02$, $d = -0.72$, two-tailed t-test), but 3-year-olds in the HA condition ($M_4 = 2.11$, $SD_4 = .63$) did not differ from chance ($p = .40$, two-tailed t-test).

Experiment 6 Discussion.

In Experiment 6, we provided children with linguistic cues that included phrasal contrast ("the chromium one, not that one") but not semantic contrast (no property word besides *chromium*). As in previous studies, we either presented children with perceptual cues that were high in alignment (HA condition) or low in alignment (LA condition).

The low-quality linguistic cues had a detrimental effect on children's indirect word learning, starting from the initial mapping between the novel color and word. We found a significant drop in 3-year-olds' accuracy during the Initial Exposure, despite the gestural cues indicating the non-target object. However, their performance on the filler trials of the Initial Exposure were still high and comparable to performance in other experiments, suggesting failure on the chromium trial was not due to confusion about the pointing task, nor to lack of engagement.

Even 3-and 4-year-olds who successfully pointed to the chromium object during the Initial Exposure did not learn the meaning of the word, as evidenced by their poor performance on the later Meaning Assessment task. Neither children in the HA nor LA condition were able to identify chromium objects at above criterion rates, despite showing false alarm rates that were at or below chance, as in previous experiments. We will discuss this in more detail in the next session, in which we compare the findings of Experiment 1, Experiment 5, and Experiment 6.

Cross Experiment Comparisons

To explore the effect of linguistic cues on preschoolers' incidental property word learning, we compared children's performance on the Meaning Assessment in Experiment 1, Experiment 5, and Experiment 6. In all three experiments, 3-and 4-year-olds were presented with either high-alignment (HA condition) or low-alignment (LA) perceptual cues. However, the experiments differed in the quality of the linguistic cues presented to children. In Experiment 1, children received high quality linguistic cues that included both semantic contrast and syntactic parallelism ("Can you point to the chromium one? The chromium one, not the blue one."). In Experiment 5, the linguistic cues included semantic contrast but not parallel syntactic structure ("I don't like the blue one. Can you point to the one that's chromium?"). In Experiment 6, the linguistic cues included neither semantic contrast nor syntactic parallelism ("Can you point to the chromium one? The chromium one, not that one."). Across experiments, all children were provided with sufficient information to allow most children to refer to the carrier object[21]. Thus, comparing across the three experiments, we can investigate the unique contributions of semantic

[21] Rates of correct pointing during the Initial Task ranged from 74% to 96%, and children who failed to infer the correct carrier object were not included in further analyses.

contrast and syntactic parallelism, as well as how these linguistic cues interact with perceptual alignment to impact indirect word learning.

Since we found distinct patterns of results for the 3-year-olds and 4-year-olds in Experiment 5 and Experiment 6, we analyzed the findings for the two age groups separately. We conducted two-way ANOVAs with Hits as the dependent variable and Alignability (HA or LA) and Linguistic Cues (Experiment 1, Experiment 5, or Experiment 6) as the independent variables for each age group.

4-year-olds' Meaning Assessment comparisons. We found a significant main effect of Alignment ($F(1,127) = 12.68$, $p = .001$, $\eta^2 = .09$) and a significant main effect of Linguistic Cues/Experiments ($F(1,127) = 3.95$, $p = .02$, $\eta^2 = .06$). The interaction between Alignment and Linguistic Cues was not significant. Further examination revealed that across experiments, 4-year-olds presented with highly alignable object pairs (HA condition, $M = 2.24$, $SD = 0.96$) correctly identified more chromium objects than children presented with objects pairs that were low in alignability (LA condition, $M = 1.58$, $SD = 1.03$). Four-year-olds made significantly more hits in Experiment 1 than Experiment 6 ($p < .05$, Bonferroni test), mainly driven by the HA group (See Figure 18). The difference between Linguistic Cues/Experiments was not significant for 4-year-olds in the LA condition. Figure 16 shows the mean number of hits made by 4-year-olds in Experiment 1, 5, and 6.

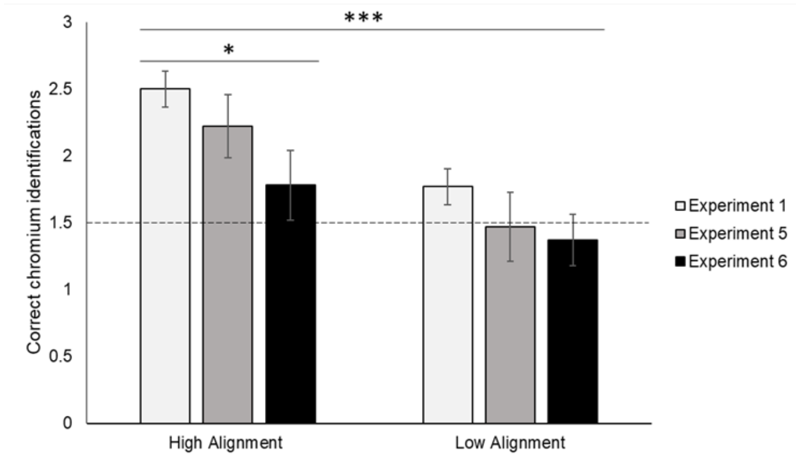


Figure 18. Mean number of chromium objects (max.= 3) correctly endorsed as *chromium* (hits) by 4-year-olds in Experiments 1, 5, and 6. Error bars depict +/- one standard error. The dashed line represents chance performance. * $p < .05$, *** $p < .005$.

3-year-olds' Meaning Assessment Comparisons. We conducted the same model for 3-year-olds as we did for 4-year-olds. We found no significant effects; however, there was a non-significant trend for children in the HA condition to make more hits than those in the LA condition ($M_{HA} = 2.10$, $SD_{HA} = 1.06$; $M_{LA} = 1.70$, $SD_{LA} = 1.01$), $F(1,121) = 3.25$, $p = .07$, $\eta^2 = .03$). Figure 19 shows the mean number of hits made by 3-year-olds in Experiment 1, 5, and 6.

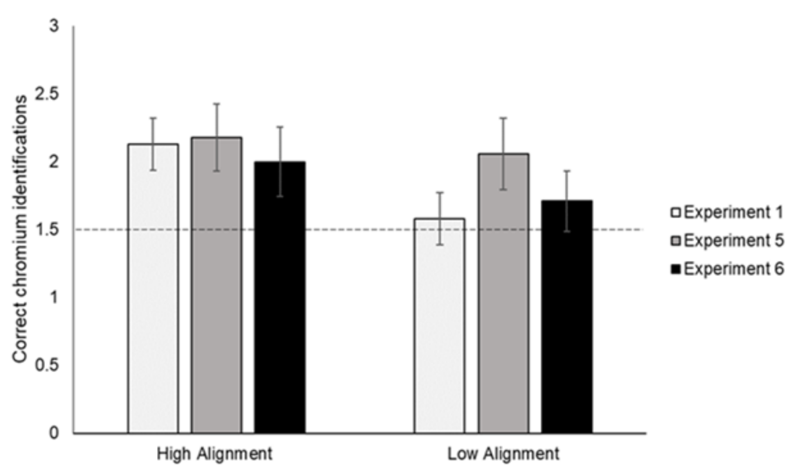


Figure 19. Mean number of chromium objects (max.= 3) correctly endorsed as *chromium* (hits) by 3-year-olds in Experiments 1, 5, and 6. Error bars depict +/- one standard error. The dashed line represents chance performance.

Chapter Discussion

Comparing across Experiments 1, 5, and 6 allowed us to examine the unique contributions of linguistic cues, as well as how linguistic cues interacted with perceptual cues. We found that 4-year-olds were sensitive to both types of cues, although the effects of the perceptual ones seemed to be stronger than those of the linguistic ones. When presented with low-quality perceptual cues (the LA condition), 4-year-olds had difficulties linking a novel color word to its referent, even when given high quality linguistic input that included semantic contrast, parallel syntax, and a clear reference to the target object (Experiment 1). However, perceptual cues were not the sole factor influencing children's indirect property-word learning. Among the 4-year-olds who were presented with high quality perceptual cues (the HA condition), there was a decline in their performances as the number of types of cues embedded within the linguistic input gradually decreased. In fact, when provided with linguistic input that included neither semantic contrast nor parallel syntax (Experiment 6), 4-year-olds in the HA condition performed at chance, just like 4-year-olds in the LA condition. Findings for the 3-year-olds paralleled those of the 4-year-olds, in that when both groups performed relatively well when presented with high-quality linguistic input and performed poorly when presented with low-quality linguistic input. However, 3-year-olds' performances differed from those of 4-year-olds when presented with intermediate-quality linguistic instructions, such as in Experiment 5.

In Experiment 5, the linguistic input included a reference to the target object and semantic contrast, but not parallel syntax. We found that, under such conditions, 4-year-olds in the HA condition performed significantly better than 4-year-olds in the LA condition, with only those in the HA condition learning the word. However, we found no differences between the performances of 3-year-olds in the HA and LA conditions. Regardless of the perceptual cues that they were given, 3-year-olds made significantly more hits than expected by chance (and in addition showed a high rate of false alarms).

Why did 3-year-olds in the HA and LA conditions perform comparably in Experiment 5? We suspect that many 3-year-olds failed to discriminate between *chromium* and non-*chromium* objects, and therefore resorted to a "yes" guessing bias in the Meaning Assessment task. We found a similar pattern of results among the 3-year-olds in Experiment 4 (Chapter 4). In Experiment 4, 3-year-olds did not seem to form a link between the novel word and color during the Initial Exposure on Day 1. On Day 2, these children—who presumably did not know the meaning of the novel word—showed an overall tendency to say "yes," resulting in high hit rates and high false alarm rates, with corresponding low *d'* scores (low discrimination rates). Although we cannot examine *d'* in Experiment 5 due to the low number of items, we also found that children in both the HA and LA conditions had relatively high false alarm rates in addition to high hit rates.

The semantic cues presented in Experiment 5 ("I don't like the blue one. Can you point to the one that's chromium?") included sufficient information for children to point to the target object even before the novel word *chromium* was introduced. Three-year-olds in Experiment 5 may have engaged in the pointing without necessarily processing the novel word or comparing the two objects. Thus, although 3-year-olds were able to correctly point to the *chromium* object during the Initial Exposure, they could not discriminate between the *chromium* and non-*chromium* objects in the later Meaning Assessment task.

The high-quality linguistic input in Experiment 1 ("the chromium one, not the blue one") used the parallel syntactic structure "the X one, not the Y one" to simultaneously refer to the target object and contrast *chromium* with *blue*. Parallel syntax not only set the stage for the other two types of linguistic cues, but also invited children to compare and contrast the object pair. We propose that spontaneous comparison is crucial for indirect word learning. Even 4-year-olds did not learn the meaning of the novel word when the Initial Exposure did not facilitate comparison and contrast, such as including hard-to-align perceptual cues (LA condition across studies) or low-quality linguistic input (Experiment 6). Furthermore, Heibeck and Markman (1987) found that when children were presented with multiple cues to compare and contrast—both high alignment perceptual cues and parallel syntactic structure—children could learn a novel color even in the absence of semantic contrast.

Taken together, the findings from our experiments and Heibeck and Markman's (1987) studies suggest that both linguistic and perceptual cues are important for promoting children's indirect property-word learning. Linguistic cues interact with perceptual cues to direct children's attention to both the novel word and potential candidates to map the novel word onto. One possible way for linguistic cues to do so is through inviting spontaneous comparison and contrast. When such processes are not elicited, children may have difficulties learning a novel property word in indirect situations.

Chapter 8: General Discussion

One of the main difficulties of word learning is determining a novel word's referent (Quine, 1960). This is especially true for property words, whose referents are not naturally partitioned into distinct entities (Gentner, 1982; 2006; 2016; Gentner & Boroditsky, 2001). Ways of increasing referential transparency have been shown to facilitate property word acquisition in direct learning situations, such as using familiar objects to demonstrate the property (e.g. Hall, Waxman, & Hurwitz, 1993; Markman & Wachtel, 1988), giving a clear indication that the novel word is a modifier of a noun (e.g. Mintz, 2005; Mintz & Gleitman, 2002), and using property-emphasizing gestures (e.g. Hall et al., 2010; O' Neill, Topolovec, & Stern-Cavalcante, 2002).

Structure-mapping processes of comparison and contrast can also promote referential transparency. We focused on two avenues for inviting spontaneous comparison and contrast: (1) visually, through manipulating the ease of perceptual alignment; and (2) verbally, through manipulating the amount of contrast embedded in the linguistic input. To investigate the potential of these methods in promoting referential transparency (and thus, word learning), we presented them in a highly challenging situation—indirect learning of property words.

Through a series of six studies, we found evidence that both high perceptual alignment and linguistic contrast can promote preschoolers' property-word learning, even in indirect situations. We briefly exposed 3- and 4-year-olds to a novel word (*chromium*) for an unfamiliar color (olive green), while manipulating the alignability of the object pair presented and/or the amount of contrast embedded in the linguistic input that children were given. Approximately ten minutes after the Initial Exposure to the word and color, we assessed how well children were able to apply the novel word to previously seen olive-green objects as well as new olive-green objects. Overall, most children successfully pointed to the *chromium* object during the Initial Exposure; however, how much the children learned from this experience differed greatly depending the perceptual and linguistic input.

We found an overall pattern for children to show better learning when presented with easy-to-align perceptual cues (High Alignment [HA] condition) than hard-to-align perceptual cues (Low Alignment [LA]). This advantage was especially strong when the high-alignment perceptual cues were presented in conjunction with high-quality linguistic cues (Chapter 2 Experiment 1). The difference between high and low perceptual alignment cannot be attributed to the quantity of information alone. Although from a hypothesis testing view, one can argue that high-alignment perceptual cues offer a more targeted hypothesis space than low-alignment perceptual cues, we still found a high alignment advantage when all children were provided with sufficient information to logically infer the correct mapping (Chapter 2 Experiment 2). In addition, children in the HA condition learned just as well in indirect learning situations as in direct learning situations (Chapter 3). Even when presented with unfamiliar shapes during the Initial Exposure, 4-year-olds in the HA condition outperformed those in the LA condition. This high perceptual alignment advantage persisted 2-4 days after the initial exposure, and to new instances (Chapter 4).

We found that linguistic cues also influenced children's indirect property-word learning. While 4-year-olds in the LA condition never demonstrated indirect property-word learning, learning by 4-year-olds in the HA condition differed depending on the linguistic input they received. Children who heard lower-quality linguistic cues during the initial exposure performed worse on a later word meaning assessment task than children who heard higher-quality linguistic cues (Chapter 7). In fact, even 4-year-olds in the HA condition failed to learn the novel word when presented with low-quality linguistic input. Findings for 3-year-olds replicated those for 4-year-olds when children were presented with linguistic cues that were either clearly of high quality (including semantic contrast, parallel syntax, and a reference to the target object) or clearly of low quality (including only a reference to the target object). However, when the linguistic cues were of intermediate quality (only semantic contrast and a reference to the target object), 3-year-olds did not learn the new word.

Overall, these findings suggest that high perceptual alignment and linguistic contrast promotes indirect property-word learning. High perceptual alignment not only invites spontaneous comparison and contrast of the available objects, but also increases the fluency of the comparison process. When the referent property is presented as an alignable difference between the two objects, the property "pops out" and offers itself as a salient candidate for the novel word's referent. Similarly, linguistic cues can also direct the learner to the target property, through methods such as inviting spontaneous comparison or contrasting the new word to a known word of the same dimension. However, when comparison and contrast is made more taxing, either perceptually (such as the unfamiliar shaped objects used in Experiment 4) or linguistically (such as the intermediate-quality linguistic cues presented in Experiment 5), children fail to learn a new property word in an indirect situation.

High quality input promotes word learning

For the past two decades, much emphasis has been placed on the strong positive correlation between the amount of language a child hears and the child's vocabulary size (e.g. Goodman, Dale, & Li, 2008; Hart & Risley, 2003; Hoff, 2003; Huttenlocher et al., 1991). More recently, researchers have shifted focus from the sheer quantity of language input to the quality of the input, such as the referential transparency of new words (e.g. Cartmill et al., 2013; Hirsh-Pasek et al., 2015; Rowe, 2012). These studies have mainly examined the effect of input quality on noun learning. Here, we present evidence suggesting that high quality input also promotes property word learning, even in indirect situations.

As discussed earlier, properties are not naturally individualized by the physical world; thus, property words are inherently low on referential transparency. Even if a learner can successfully link a novel property word to a correct referent object, she would still need to select which particular property the word refers to. Indeed, as we found across experiments, even when children pinpoint a correct referent object (during the Initial Exposure), they do not necessarily map the word onto the referent property (as indicated by children in the LA conditions). Highlighting the target property, not just an object exemplifying the property, is crucial for increasing the referential transparency of property words.

Situations that involve comparison and contrast of perceptually aligned objects are especially effective for emphasizing specific properties. Waxman and Klibanoff (2000) found that the quality of the objects presented was crucial to whether children were able to learn a new word in a direct learning situation. Many other researchers, although not explicitly testing the effects of comparison and contrast, have also intuitively incorporated comparison and contrast in their experimental designs.

Here, we present the first empirical evidence suggesting that high quality input can facilitate indirect property word learning. When children were given high quality exposures (high perceptual alignment and high linguistic support), they learned a new property word despite briefness and indirectness of the exposures.

Of course, we are not suggesting that overall quantity of exposure is not important. In fact, we found that children presented with hard-to-align objects generally benefited from additional exposure (Chapter 7). We speculate that additional exposure can benefit incidental word learning in at least two ways. First, with more exposure, the probability that one would encounter high quality individual instances increases as well. This does not apply to our controlled experimental manipulations but may very likely occur in real life. Second, more exposure—especially over a relatively short period—may elicit comparison *across* individual instances. This might explain children's overall elevated performance on Day 2 of Experiment 4. Seeing the Retention task (which was identical to Day 1's Meaning Assessment task) may have elicited retrieval of Day 1's events, as well as comparison of Day 1's events to the task at hand.

Thus, even though children were not given corrective feedback on Day 1, their performances improved on Day 2.

Indirect word learning in real life likely includes relatively few high-quality learning opportunities dispersed among mostly low-quality potential learning situations across long periods. In some lucky cases, the initial word-meaning mapping formed during a high-quality situation is rich and robust. However, in many other cases, the representation still needs to be enriched and consolidated through a prolonged process.

Fast mapping and slow learning

In their classic 1978 indirect property word learning study, Carey and Bartlett coined the term fast-mapping—the ability to quickly link a novel word to its referent with limited exposure. Although approximately half of the children in the experiment demonstrated some learning of the novel word and color when tested later, the other half did not seem to retain any learning. Since then, many studies have also found a disconnect between children's fast mapping and slow learning (e.g. Bion, Borovsky, & Fernald, 2013; Horst & Samuelson, 2008; Wagner, Dobkins, & Barner, 2013).

Most of the fast-mapping literature has focused on children's abilities to map a noun label to a novel object (e.g., Golinkoff et al., 1992; Schafer & Plunkett, 1998; Spiegel & Halberda, 2012; Wilkinson & Mazzitelli, 2003). In such cases, children are presented with a novel object and at least one familiar object upon hearing a novel label. Children are generally very accurate at picking out the correct object under such circumstances. However, there are mixed findings on whether children can retain this performance. For example, Horst and Samuelson found that, although 2-year-olds were able to correctly select the novel object among known objects upon hearing a novel word, they were at chance merely five minutes later.

McMurray, Horst, Samuelson, and colleagues have proposed that in-the-moment referent selection and long-term word learning are independent processes with different mechanisms (Horst & Samuelson, 2008; Horst, & Samuelson, 2012; McMurray, Horst, Toscano, & Samuelson, 2009). They suggest that online referent determination can be achieved through dynamic competition between referents and/or words. The goal for referent selection is "simply arriving at a state in which one word and one object are under consideration" (McMurray, Horst, & Samuelson, 2012). Forming a word-object linkage is not required, and if a linkage is formed, it does not need to be complete nor committed to. Thus, children may display highly competent behavior despite poor representations and little learning. The authors also acknowledge that referent selection and learning, though logically distinct, can be "deeply and subtly related."

Our studies provided evidence supporting the distinction between selecting a referent object and learning a new word. Across experiments, most children in the LA condition passed the Initial Exposure pointing task—that is, they pointed to the target object when asked to "point to the chromium one, not the blue one." However, high performance on the Initial Exposure did not necessarily transfer to high performance on the later Meaning Assessment task. This disassociation suggest that children did not need to learn the meaning of the novel word *chromium* to select the correct referent object. We propose that children could have correctly selected the target object in the Initial Exposure by simply responding to the linguistic cue "not the blue one." They did not need to determine the target property dimension (color), let alone the value along that dimension (olive green). In other words, children could select the correct referent object without learning the meaning of the word.

At the same time, our studies suggest one way that referent selection and word learning may be related—when the referential transparency of a given situation is extremely high (such as the high perceptual alignment and informative linguistic contrast situations we created), a link between the word and its referent is formed spontaneously. Under such circumstances, fast-mapping can also lead to fast-learning.

Conclusion

To the best of our knowledge, this is the first set of studies to explicitly test the role of structural alignment in indirect word learning situations. Much prior work has shown that structural alignment processes support children's learning in explicit learning tasks, such as mathematics learning (Richland, Zur, & Holyaok, 2007; Richland, Holyoak, & Stigler. 2004; Rittle-Johnson & Star, 2007; Thompson & Opfer, 2010), word learning (e.g., Gentner, Loewenstein, & Hung, 2007; Graham et al., 2010), and science education (Jee et al., 2010; Kurtz & Gentner, 2013). For the first time we have shown that structural alignment processes promote learning even when learning was not the focus of the task.

We argue that, when presented with high-quality input, children can quickly and effortlessly determine a likely referential candidate for a novel adjective without engaging in deliberate reasoning. Indeed, in Experiment 2 and Experiment 4, children in the LA condition performed worse than those in the HA condition in the Meaning Assessment task, despite receiving sufficient information across two trials. This suggests that cross-situational hypothesis testing was not the main mechanism employed by our participants. Due to the affordances of the initial exposure, children in the HA condition were unlikely to have been provoked to use other strategies. We do not think that children in the LA condition were engaged in deliberate hypothesis-testing either, since they performed better when explicitly told that they had to learn the meaning of the word *chromium* than when they were not provided with this instruction (Experiment 3). Without explicit instructions, children did not seem to be focused on deciphering the meaning of *chromium*. As previously discussed, it is more likely that they were just trying to comply with the experimenter's pointing request. Of course, we are not claiming that children are not capable of learning words through this technique; much research has suggested that adults, children, and even infants can form word-meaning mappings though accumulating evidence over time (e.g. Siskind, 1996; Smith & Yu, 2008; Trueswell, Medina, Hafri, & Gleitman, 2013; Yu & Smith, 2007; Xu & Tenenbaum, 2007). However, due to the construction of our Initial Exposure task, that does not seem to be the case for the participants in our studies.

Incidental learning experiences are clearly an important part of children's language learning, but they are almost invisible in nature. Such experiences can occur at any time and place, with no necessary learning intention of the knowledge receiver or pedagogical intention of the knowledge provider. Yet during serendipitous moments, when the receiver is faced with high-quality input, he or she can notice, and even acquire, new information without external guidance. Recent research has found that quality of input, rather than overall quantity of input, is a strong predictor of children's vocabulary growth (e.g. Cartmill et al., 2013; Hirsh-Pasek et al., 2015; Rowe, 2012). In this paper we show that comparison and contrast are powerful tools for creating high quality input for spontaneous property word learning. When the target information is highlighted through comparison processes, it is more likely that spontaneous learning will take place.

References

Akhtar, N. (2005). The robustness of learning through overhearing. *Developmental Science*, *8*(2), 199-209.

Akhtar, N., Jipson, J., & Callanan, M. A. (2001). Learning words through overhearing. *Child Development, 72*(2), 416-430.

Au, T. K. F. (1990). Children's use of information in word learning. *Journal of Child Language*, *17*(2), 393-416.

Au, T. K. F., & Laframboise, D. E. (1990). Acquiring color names via linguistic contrast: The influence of contrasting terms. *Child Development*, *61*(6), 1808-1823.

Au, T. K. F., & Markman, E. M. (1987). Acquiring word meanings via linguistic contrast. *Cognitive Development*, *2*(3), 217-236.

Bates, E., Marchman, V., Thal, D., Fenson, L., Dale, P., Reznick, J. S., ... & Hartung, J. (1994). Developmental and stylistic variation in the composition of early vocabulary. *Journal of Child Language*, *21*(1), 85-123.

Baldwin, D. A., Markman, E. M., Bill, B., Desjardins, R. N., Irwin, J. M., & Tidball, G. (1996). Infants' reliance on a social criterion for establishing word-object relations. *Child Development, 67(6),* 3135-3153.

Behne, T., Carpenter, M., & Tomasello, M. (2005). One-year-olds comprehend the communicative intentions behind gestures in a hiding game. *Developmental Science, 8(6)*, 492-499.

Bion, R. A., Borovsky, A., & Fernald, A. (2013). Fast mapping, slow learning: Disambiguation of novel word–object mappings in relation to vocabulary learning at 18, 24, and 30 months. *Cognition, 126*(1), 39-53.

Blewitt, P., Rump, K. M., Shealy, S. E., & Cook, S. A. (2009). Shared book reading: When and how questions affect young children's word learning. *Journal of Educational Psychology, 101(2),* 294.

Bloom, P., & Kelemen, D. (1995). Syntactic cues in the acquisition of collective nouns. *Cognition, 56(1),* 1-30.

Bornstein, M. H., Cote, L. R., Maital, S., Painter, K., Park, S. Y., Pascual, L., ... & Vyt, A. (2004). Cross-Linguistic Analysis of Vocabulary in Young Children: Spanish, Dutch, French, Hebrew, Italian, Korean, and American English. *Child Development, 75(4),* 1115-1139.

Booth, A. E., & Waxman, S. (2002). Object names and object functions serve as cues to categories for infants. *Developmental Psychology*, *38*(6), 948.

Booth, A. E., & Waxman, S. R. (2003). Mapping words to the world in infancy: Infants' expectations for count nouns and adjectives. *Journal of Cognition and Development*, *4*(3), 357-381.

Booth, A. E., & Waxman, S. R. (2009). A horse of a different color: Specifying with precision infants' mappings of novel nouns and adjectives. *Child Development*, *80*(1), 15-22.

Braginsky, M., Yurovsky, D., Marchman, V. A., & Frank, M. C. (2015). Developmental changes in the relationship between grammar and the lexicon. In D. C. Noelle, R. Dale, A. S. Warlaumont, J. Yoshimi, T. Matlock, C. D. Jennings, & P. P. Maglio (Eds.), *Proceedings of the 37th Annual Conference of the Cognitive Science Society* (pp. 1 – 6). Austin, TX: Cognitive Science Society.

Brooks, R., & Meltzoff, A. N. (2008). Infant gaze following and pointing predict accelerated vocabulary growth through two years of age: A longitudinal, growth curve modeling study. *Journal of Child Language*, *35*(1), 207-220.

Callanan, M. A., Akhtar, N., & Sussman, L. (2014). Learning words from labeling and directive speech. *First Language*, *34*(5), 450-461.

Carey, S., & Bartlett, E. (1978). Acquiring a single new word. *Proceedings of the Stanford Child Language Conference, 15,* 17-29.

Cartmill, E. A., Armstrong, B. F., Gleitman, L. R., Goldin-Meadow, S., Medina, T. N., & Trueswell, J. C. (2013). Quality of early parent input predicts child vocabulary 3 years later. *Proceedings of the National Academy of Sciences*, *110*(28), 11278-11283.

Childers, J. B., & Paik, J. H. (2009). Korean-and English-speaking children use cross-situational information to learn novel predicate terms. *Journal of Child Language*, *36*(1), 201-224.

Clark, E. V. (1987). The principle of contrast: A constraint on language acquisition. In B. J. MacWhinney (Ed.), *Mechanisms of language acquisition,* (pp.1-33). Hover, England: Psychology Press.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods,* 39, 175-191.

Ferguson, B., & Lew-Williams, C. (2016). Communicative signals support abstract rule learning by 7-month-old infants. *Scientific Reports*, *6*, 25434.

Fernald, A., Zangl, R., Portillo, A. L., & Marchman, V. A. (2008). Looking while listening: Using eye movements to monitor spoken language. *Developmental psycholinguistics: On-line methods in children's language processing*, *44*, 97.

Fernald, A., & Mazzie, C. (1991). Prosody and focus in speech to infants and adults. *Developmental Psychology*, *27*(2), 209.

Fisher, C. (1996). Structural limits on verb mapping: The role of analogy in children's interpretations of sentences. *Cognitive Psychology*, *31*(1), 41-81.

Fisher, C., Gertner, Y., Scott, R. M., & Yuan, S. (2010). Syntactic bootstrapping. *Wiley Interdisciplinary Reviews: Cognitive Science, 1(2),* 143-149.

Gampe, A., Liebal, K., & Tomasello, M. (2012). Eighteen-month-olds learn novel words through overhearing. *First Language*, *32*(3), 385-397.

Gasser, M., & Smith, L. B. (1998). Learning nouns and adjectives: A connectionist account. *Language and Cognitive Processes*, *13*(2-3), 269-306.

Gentner, D. (1982). Why nouns are learned before verbs: Linguistic relativity versus natural partitioning. In S. A. Kuczaj (Ed.), *Language development: Vol. 2. Language, thought and culture* (pp. 301-334). Hillsdale, NJ: Lawrence Erlbaum Associates.

Foster, E. K., & Hund, A. M. (2012). The impact of scaffolding and overhearing on young children's use of the spatial terms between and middle. *Journal of Child Language*, *39*(2), 338-364.

Gentner, D. (1982). Why nouns are learned before verbs: Linguistic relativity versus natural partitioning. In S. A. Kuczaj (Ed.), *Language development: Vol. 2. Language thought and culture* (pp. 301-334). Hillsdale, NJ: Lawrence Erlbaum Associates.

Gentner, D. (1983). Structure-Mapping: A theoretical framework for analogy. *Cognitive Science,* *7*(2), 155-170.

Gentner, D. (2003). Why we're so smart. In D. Gentner and S. Goldin-Meadow (Eds.), *Language in mind: Advances in the study of language and thought* (pp.195-235). Cambridge, MA: MIT Press.

Gentner, D. (2006). Why verbs are hard to learn. In K. Hirsh-Pasek, & R. Golinkoff, (Eds.) *Action meets word: How children learn verbs* (pp. 544-564). Oxford University Press.

Gentner, D. (2010). Bootstrapping the mind: Analogical processes and symbol systems. *Cognitive Science, 34*(5), 752-775.

Gentner, D. (2016). Language as cognitive toolkit: How language supports relational thought. *American Psychologist. 71(8),* 650-657.

Gentner, D., & Boroditsky, L. (2001). Individuation, relativity and early word learning. In M. Bowerman & S. Levinson (Eds.), *Language acquisition and conceptual development* (pp. 215-256). Cambridge, UK: Cambridge University Press.

Gentner, D., & Boroditsky, L. (2009). Early acquisition of nouns and verbs: Evidence from Navajo. In V. Gathercole (Ed.) *Routes to Language: Studies in honor of Melissa Bowerman,* 5-36. Taylor & Francis, New York, NY.

Gentner, D., & Gunn, V. (2001). Structural alignment facilitates the noticing of differences. *Memory and Cognition, 29*(4), 565-577.

Gentner, D., & Hoyos, C. (2017). Analogy and abstraction. *Topics in Cognitive Science*, 9(3), 672-693.

Gentner, D., Loewenstein, J., & Hung, B. (2007). Comparison facilitates children's learning of names for parts. *Journal of Cognition and Development, 8*(3), 285-307.

Gentner, D., & Markman, A. B. (1994). Structural alignment in comparison: No difference without similarity. *Psychological Science, 5*(3), 152-158.

Gleitman, L. (1990). The structural sources of verb meanings. *Language Acquisition*, *1*(1), 3-55.

Golinkoff, R. M., Hirsh-Pasek, K., Bailey, L. M., & Wenger, N. R. (1992). Young children and adults use lexical principles to learn new nouns. *Developmental Psychology, 28*(1), 99

Goodman, J. C., Dale, P. S., & Li, P. (2008). Does frequency count? Parental input and the acquisition of vocabulary. *Journal of child language*, 35(3), 515-531.

Graham, S., A., Namy, L. L., Gentner, D., & Meagher, K. (2010). The role of comparison in preschoolers' novel object categorization. *Journal of Experimental Child Psychology, 107*. 280-290.

Hall, D. G., Williams, S. G., & Bélanger, J. (2010). Learning count nouns and adjectives: Understanding the contributions of lexical form class and social-pragmatic cues. *Journal of Cognition and Development*, *11*(1), 86-120.

Hall, D. G., Waxman, S. R., & Hurwitz, W. M. (1993). How two-and four-year-old children interpret adjectives and count nouns. *Child Development*, *64*(6), 1651-1664.

Hart, B., & Risley, T. R. (2003). The early catastrophe: The 30 million word gap by age 3. *American Educator*, *27*(1), 4-9.

Heibeck, T. H., & Markman, E. M. (1987). Word learning in children: An examination of fast mapping. *Child Development,* 1021-1034.

Haryu, E., Imai, M., & Okada, H. (2011). Object similarity bootstraps young children to action-based verb extension. Child Development, 82(2), 674-686.Jee, B. D., Uttal, D. H., Gentner, D., Manduca, C., Shipley, T., Sageman, B., Ormand, C. J., & Tikoff, B. (2010). Analogical thinking in geoscience education. *Journal of Geoscience Education*, *58* (1), 2-13A

Hirsh-Pasek, K., Adamson, L. B., Bakeman, R., Owen, M. T., Golinkoff, R. M., Pace, A., ... & Suma, K. (2015). The contribution of early communication quality to low-income children's language success. *Psychological Science*, *26*(7), 1071-1083.

Hoff, E. (2003). The specificity of environmental influence: Socioeconomic status affects early vocabulary development via maternal speech. *Child Development,* 74(5), 1368-1378.

Hoff, E. (2006). How social contexts support and shape language development. *Developmental Review*, *26*(1), 55-88.

Horst, J. S., & Samuelson, L. K. (2008). Fast mapping but poor retention by 24-month-old infants. *Infancy*, *13*(2), 128-157.

Huttenlocher, J., Haight, W., Bryk, A., Seltzer, M., & Lyons, T. (1991). Early vocabulary growth: Relation to language input and gender. *Developmental Psychology,* 27(2), 236.

Jee, B. D., Uttal, D. H., Gentner, D., Manduca, C., Shipley, T. F., Tikoff, B., ... & Sageman, B. (2010). Commentary: Analogical thinking in geoscience education. *Journal of Geoscience Education*, *58*(1), 2-13.

Jusczyk, P. W., Houston, D. M., & Newsome, M. (1999). The beginnings of word segmentation in English-learning infants. *Cognitive psychology*, *39*(3-4), 159-207.

Justice, L. M. (2002). Word exposure conditions and preschoolers' novel word learning during shared storybook reading. *Reading Psychology, 23(2),* 87-106

Krcmar, M., Grela, B., & Lin, K. (2007). Can toddlers learn vocabulary from television? An experimental approach. *Media Psychology, 10*(1), 41-63.

Kurtz, K. J., & Gentner, D. (2013). Detecting anomalous features in complex stimuli: the role of structured comparison. *Journal of Experimental Psychology: Applied*, *19*(3), 219.

Landau, B., Smith, L. B., & Jones, S. S. (1988). The importance of shape in early lexical learning. *Cognitive Development*, *3*(3), 299-321.

Markman, A. B., & Gentner, D. (1993). Structural alignment during similarity comparisons. *Cognitive Psychology, 25*(4), 431-467.

Markman, E. M. (1989). *Categorization and naming in children: Problems of induction.* Cambridge, MA: MIT Press.

Markman, E. M., & Wachtel, G. F. (1988). Children's use of mutual exclusivity to constrain the meanings of words. *Cognitive Psychology*, *20*(2), 121-157.

Mattys, S. L., Jusczyk, P. W., Luce, P. A., & Morgan, J. L. (1999). Phonotactic and prosodic effects on word segmentation in infants. *Cognitive Psychology*, *38*(4), 465-494.

McMurray, B., Horst, J. S., & Samuelson, L. K. (2012). Word learning emerges from the interaction of online referent selection and slow associative learning. *Psychological Review*, *119*(4), 831.

Mintz, T. H. (2005). Linguistic and conceptual influences on adjective acquisition in 24-and 36-month-olds. *Developmental Psychology*, *41*(1), 17.

Mintz, T. H., & Gleitman, L. R. (2002). Adjectives really do modify nouns: The incremental and restricted nature of early adjective acquisition. *Cognition*, *84*(3), 267-293.

O'Neill, D. K., Topolovec, J., & Stern-Cavalcante, W. (2002). Feeling sponginess: The importance of descriptive gestures in 2-and 3-year-old children's acquisition of adjectives. *Journal of Cognition and Development*, *3*(3), 243-277.

Pruden, S. M., Hirsh-Pasek, K., Golinkoff, R. M., & Hennon, E. A. (2006). The birth of words: Ten-month-olds learn words through perceptual salience. *Child Development*, *77*(2), 266-280.

Quine, W. V. O. (2013). *Word and object*. Cambridge, Massachusetts: MIT press.

Richland, L. E., Holyoak, K. J., & Stigler, J. W. (2004). Analogy use in eighth-grade mathematics classrooms. *Cognitiion and Instruction, 22*, 37-60.

Richland, L. E., Zur, O., & Holyoak, K. J. (2007). Cognitive supports for analogies in the mathematics classroom. *Science, 316*, 1128-1129.

Rittle-Johnson, B., & Star, J. R. (2007). Does comparing solution methods facilitate conceptual and procedural knowledge? An experimental study on learning to solve equations. *Journal of Educational Psychology*, *99*, 561–574.

Roseberry, S., Hirsh-Pasek, K., Parish-Morris, J., & Golinkoff, R. M. (2009). Live action: Can young children learn verbs from video?. *Child Development, 80(5),* 1360-1375.

Rowe, M. L. (2012). A longitudinal investigation of the role of quantity and quality of child-directed speech in vocabulary development. *Child Development*, *83*(5), 1762-1774.

Sagi, E., Gentner, D., & Lovett, A. (2012). What difference reveals about similarity. *Cognitive Science, 36*(6), 1019-1050.

Sandhofer, C. M., & Smith, L. B. (1999). Learning color words involves learning a system of mappings.

Developmental Psychology, 35(3), 668.

Salerni, N., Assanelli, A., D'Odorico, L., & Rossi, G. (2007). Qualitative aspects of productive vocabulary at the 200-and 500-word stages: A comparison between spontaneous speech and parental

report data. *First Language*, *27*(1), 75-87.

Schafer, G., & Plunkett, K. (1998). Rapid word learning by fifteen-month-olds under tightly controlled conditions. *Child Development,* 69(2), 309-320.

Sénéchal, M., Thomas, E., & Monker, J. A. (1995). Individual differences in 4-year-old children's acquisition of vocabulary during storybook reading. *Journal of Educational Psychology, 87(2),* 218.

Siskind, J. M. (1996). A computational study of cross-situational techniques for learning word-to-meaning mappings. *Cognition, 61(1)*, 39-91.

Spiegel, C., & Halberda, J. (2011). Rapid fast-mapping abilities in 2-year-olds. *Journal of Experimental Child Psychology*, *109*(1), 132-140.

Smith, L., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition, 106(3),* 1558-1568.

Taylor, M., & Gelman, S. A. (1988). Adjectives and nouns: Children's strategies for learning new words. *Child Development*, 411-419.

Thompson, C. A., & Opfer, J. E. (2010). How 15 hundred is like 15 cherries: Effect of progressive alignment on representational changes in numerical cognition. *Child Development, 81*(6), 1768-1786.

Tomasello, M., & Barton, M. E. (1994). Learning words in non-ostensive contexts. *Developmental Psychology, 30(5)*, 639.

Tomasello, M., & Farrar, M. J. (1986). Joint attention and early language. *Child Development*, 1454-1463.

Tribushinina, E., Van Den Bergh, H., Kilani-Schoch, M., Aksu-Koç, A., Dabašinskienė, I., Hrzica, G., ... & Dressler, W. (2013). The role of explicit contrast in adjective acquisition: A cross-linguistic longitudinal study of adjective production in spontaneous child speech and parental input. *First Language*, *33*(6), 594-616.

Trueswell, J. C., Medina, T. N., Hafri, A., & Gleitman, L. R. (2013). Propose but verify: Fast mapping meets cross-situational word learning. *Cognitive Psychology, 66(1),* 126-156.

Wagner, K., Dobkins, K., & Barner, D. (2013). Slow mapping: color word learning as a gradual inductive process. *Cognition, 127*(3), 307-317.

Waxman, S. R., & Booth, A. E. (2001). Seeing pink elephants: Fourteen-month-olds'

interpretations of

novel nouns and adjectives. *Cognitive Psychology*, *43*(3), 217-242.

Waxman, S. R., & Klibanoff, R. S. (2000). The role of comparison in the extension of novel adjectives. *Developmental Psychology, 36*(5), 571.

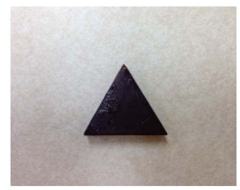
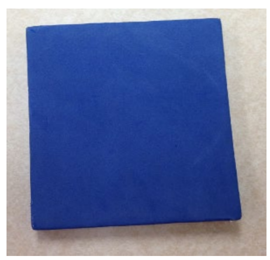
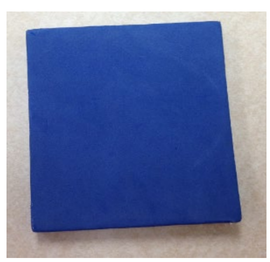
Wellwood, A., Gagliardi, A., & Lidz, J. (2016). Syntactic and lexical inference in the acquisition of novel superlatives. *Language Learning and Development*, *12*(3), 262-279.

Wilkinson, K. M., & Mazzitelli, K. (2003). The effect of ‘missing’information on children's retention of fast-mapped labels. *Journal of Child Language*, 30(1), 47-73.

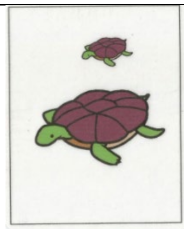
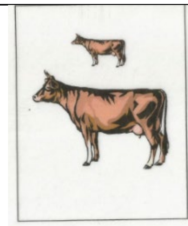
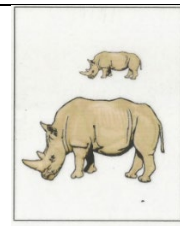
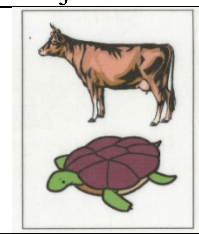
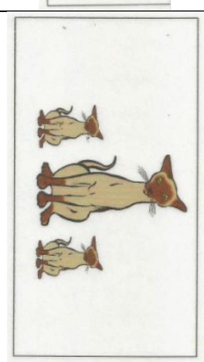
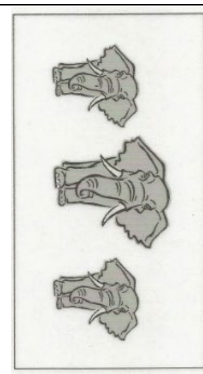
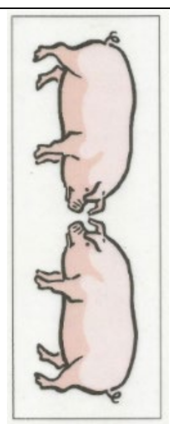
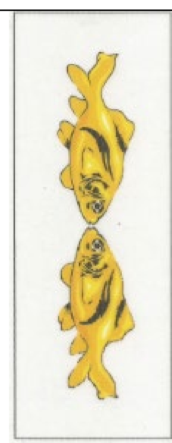
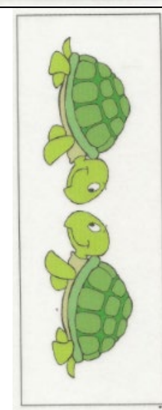
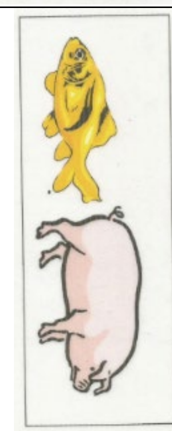
Xu, F., & Tenenbaum, J. B. (2007). Word learning as Bayesian inference. *Psychological Review,* 114(2), 245.

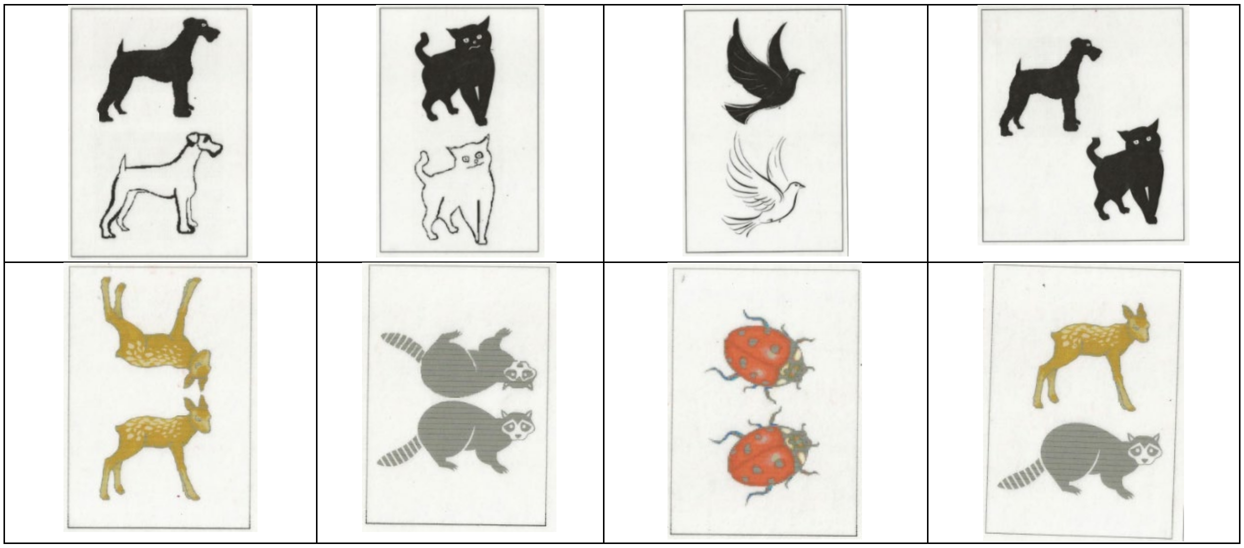
Yu, C., & Smith, L. B. (2007). Rapid word learning under uncertainty via cross-situational statistics. *Psychological Science,* 18(5), 414-420.

Appendix A. Materials used in the Initial Referent Selection Task.

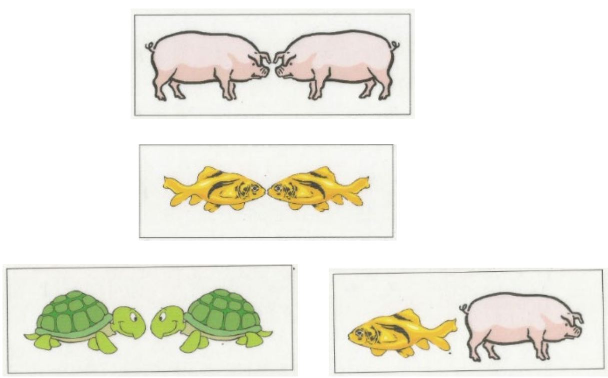
	Referent 1	Referent 2
<u>Animal Trial</u> *“Look at these two! Can you point to the cow (lion)? The cow (lion), not the lion (cow).”*		
<u>Size Trial</u> *“Look at these two! Can you point to the big (small) one? The big (small) one, not the small (big) one.”*		
<u>HA Referent-selection Trial</u> *“Look at these two! Can you point to the chromium one? The chromium one, not the blue one.”*		
<u>LA Referent-selection Trial</u> *“Look at these two! Can you point to the chromium one? The chromium one, not the blue one.”*		
<u>Length Trial</u> *“Look at these two! Can you point to the long (short) one? The long (short) one, not the short (long) one.”*		

Appendix B: Materials used in the Relational-Match-to-Sample Task.

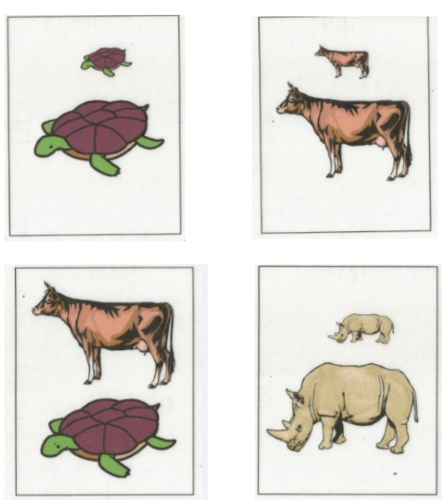
Standard 1	Standard 2	Relational Match	Object Match
			
			
			
			



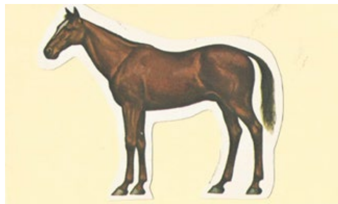
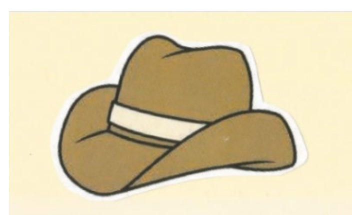
Placement 1:



Placement 2:



Appendix C. Materials used in the Yes-No Sorting Task.

	Yes Responses	No Responses
<u>Animal Set</u> *"Look at this one! Is this an animal?"*	 	 
<u>Chair Set</u> *"Look at this one! Is this a chair?"*	 	 
<u>Soft Set</u> *"Look at this one! Is this a soft one?"*