Slocum, Jeremy Y., Ph.D., December 2019 PSYCHOLOGICAL SCIENCES

THE ROLE OF METACOGNITION IN CHILDREN'S DISAMBIGUATION OF NOVEL NAME REFERENCE (136 pp.)

Dissertation Advisor: William Merriman

When shown a familiar and a novel object and asked to pick the referent of a novel label, even one-year-olds tend to favor the novel object (Halberda, 2003; Mervis & Bertrand, 1994). However, this so-called *disambiguation effect* becomes stronger as children develop through preschool age (Lewis & Frank, 2015). Advances in metacognition may play a role in this developmental trend. Preschoolers' awareness of their own lexical knowledge is associated with the strength of the disambiguation effect (Merriman & Schuster, 1991; Merriman & Bowman, 1989; Wall, Merriman, & Scofield, 2015). It is also associated with whether children can solve purely metacognitive forms of the disambiguation problem (Slocum & Merriman, 2018; Henning & Merriman, 2019). The current experiments tested the hypothesis that as the number of choices in a disambiguation problem increases, the frequency of correct response declines more sharply for children who lack awareness of lexical knowledge than for children who possessed it. The results of the first two experiments supported the main hypothesis. Two experiments also showed that awareness of lexical knowledge was associated with a more gradual increase in latency of correct solutions as number of choices increased. In Experiment 3, children's eye movements were recorded as they attempted to solve 3-, 4-, 5-, and 6-choice problems. Various aspects of children's eye movements were analyzed, including the number of familiar object foils checked, the number of revisits to the target, and the proportion of looking time spent on the target object. The current experiments advance our insight into why the "awareness-of-knowledge advantage" in solving disambiguation problems tends to increase as number of choices increases.

THE ROLE OF METACOGNITION IN CHILDREN'S DISAMBIGUATION OF NOVEL NAME REFERENCE

A dissertation submitted

to Kent State University in partial fulfillment

of the requirements for the

degree of Doctor of Philosophy

by

Jeremy Y. Slocum

December 2019

Dissertation written by

Jeremy Y. Slocum

B.A., Pennsylvania State University, 2014

M.A., Kent State University, 2017

Ph.D., Kent State University, 2019

Approved by

William Merriman, Chair, Doctoral Dissertation Committee

Clarissa Thompson, Members, Doctoral Dissertation Committee

Jeffrey Ciesla

Bradley Morris

Sarah Rilling

Accepted by

Maria Zaragoza, Chair, Department of Psychological Sciences

James Blank, Dean, College of Arts and Sciences

CHAPTERS

III. Experiment 1

IV. Experiment 2

V. Experiment 3

APPENDICES

LIST OF FIGURES

LIST OF TABLES

Acknowledgments

I want to thank my advisor, Dr. William E. Merriman, for all of his help and support in the design, preparation, and analysis of the experiments in this dissertation. Your wisdom and guidance have made more of an impact on me than you can know. I would also like to thank my fellow lab mates, Kyle Henning and Dr. Jenna Wall, for their support and guidance throughout the process. In addition, I would like to thank all of the preschools and daycares that allowed me to recruit children from their facilities in order to collect these data. Finally, I want to thank the members of my dissertation committee for providing helpful feedback of the presentation of my thesis. Thank you, Dr. Clarissa Thompson, Dr. Jeff Ciesla, Dr. Brad Morris, and Dr. Sarah Rilling.

Chapter 1. Introduction

Young children frequently face the challenge of determining the reference of novel words. While children can utilize a speaker's gaze and gesture in order to help make this determination (Baldwin, 1993; Gilga & Csibra, 2009), research has shown that even when these cues are present, children may not always detect them or find them to be precise enough to choose the correct referent. Yurovsky, Smith, and Yu (2013, Exp. 1) found that in approximately half of the instances in which a parent named an object while playing with their toddler, the reference of the name was highly ambiguous.

Children often resolve such ambiguities by favoring unfamiliar kinds of objects over objects for which they already know labels. Even children as young as 16 months old have shown this so-called *disambiguation effect* (Halberda, 2003; Markman, Wasow, & Hansen, 2003; Mervis & Bertrand, 1994). Although this strategy is not foolproof – the intended referent could be an object for which the child already has a different label – it is correct more often than not, especially when no other cues to a word's intended referent are present (Markman, 1984; Merriman, 1986). Because children commonly encounter situations in which the reference of a novel label is not clear, this strategy may play an important role in their word learning.

Not only do typically-developing children show the disambiguation effect, but so do children from a variety of special populations. These include children who are deaf or hard-of-hearing (Lederberg & Spencer, 2008), late talkers (Choi & Hwang, 2014; Mervis & Bertrand, 1995a), children with SLI (Specific Language Impairment) (Beverly & Estis, 2003, Estis & Beverly, 2015), bilingual and multilingual children (Byers-Heinlein & Werker, 2009, 2013; Davidson, Jergovic, Imami, & Theodos, 1997; Houston-Price, Caloghiris, & Raviglione, 2010), children with ASD (autism spectrum disorder) or who are at risk for ASD (Bedford et al., 2013; de Marchena et al., 2011; Preissler & Carey, 2005), and children with various forms of mental retardation (Mervis, 1995b; Romski et al., 1996; Wilkinson, & Green, 1998).

It is important to explore why children show the disambiguation effect. If we know the mechanisms underlying the effect, we may be able to influence children's mapping of novel labels by manipulating those mechanisms. This dissertation will focus on the potential role of metacognitive processes in the effect.

A metacognitive process is one that monitors or reflects on one's own knowledge or thinking (Flavell, 1979; Nelson & Narens, 1990). One example is meta-attention, which refers to knowledge about one's own attentional ability. Another is metamemory, which includes both knowledge of one's memory ability and the processes involved in monitoring memory. Yet another metacognitive process involves regulation or control, which occurs when one uses declarative metacognitive knowledge to alter behavior. The particular metacognitive processes that have been hypothesized to influence the disambiguation effect are acts of representing the novel name as an unknown label, the novel object as an unknown kind, and the familiar object as a known kind that already has a known label (Slocum & Merriman, 2018; Henning & Merriman, 2019). The hypothesis is that children show a more robust disambiguation effect whenever they represent each element of the disambiguation problem as a known or unknown element. Such representations promote their selecting the unknown kind of object over the known kind of object as the referent of the unknown label.

In this chapter, I will review current accounts of the disambiguation effect, then summarize developmental research on child metacognitive processes that could possibly influence the effect. I will then review the results of two recent investigations (Slocum & Merriman, 2018; Henning & Merriman, 2019) demonstrating that some preschool-age children are able to solve purely metacognitive versions of the disambiguation problem. These problems were structured so that they could not be solved by comparing phonological representations of labels or by comparing the activation of objects, but could be solved by comparing metacognitive representations. In Chapter 2, I will provide an overview of three experiments I conducted to further explore the possible role of metacognitive processes in the disambiguation effect. I hypothesized that as the task became more demanding, children who possess an awareness of their lexical knowledge would show an increasing advantage over children without this kind of awareness.

Current Accounts of the Disambiguation Effect

Researchers have proposed several explanations for the disambiguation effect. The three leading accounts are: Mutual Exclusivity; Pragmatic Contrast; and Competitive Activation.

Mutual Exclusivity. According to the Mutual Exclusivity account, children operate under the default assumption that the extension of one label will not overlap with the extension of another, that is, that two labels will not have exemplars in common (Markman & Wachtel, 1989; Merriman & Marazita, 1995). For example, a child who hears a novel label (e.g., “zav”) used in such a way that it could be referring to a familiar, nameable object (e.g., cup) or an unfamiliar, as-yet-unnamable object (e.g., garlic press), will choose the unfamiliar object as the intended referent after rejecting the familiar object because it already has a known label.

Mutual exclusivity is considered a default assumption rather than an ironclad rule. Children will overwrite the assumption and interpret a novel word as having exemplars in common with a familiar word if they receive sufficient evidence for such overlap. For example, 2- and 4-year-olds will select a familiar object rather than an unfamiliar one if the adult requesting the referent of the novel name points at the familiar object while looking back and forth between it and the child (Grassmann & Tomasello, 2010). That is, the disambiguation effect requires that the speaker’s act of reference be ambiguous. Note that in studies where the speaker has used a more subtle nonverbal cue to referential intent (e.g., merely staring at the familiar object), young children have tended to select the unfamiliar object (Graham, Nilsen, Collins, & Olineck, 2010; Jaswal & Hansen, 2006).

Children will accept a violation of mutual exclusivity if preserving it requires them to reject a strongly-held belief. In Merriman and Bowman (1989, Experiment 2), 2- and 3-year-olds heard a novel name used for an exemplar of a known name (e.g., *car*). Nearly all the participants accepted both names for the object if the object was a typical exemplar (e.g., a typical-looking car). However, if the object was an atypical exemplar (e.g., a hybrid of a car and truck), about half of the participants rejected the known name for it. (None did so in a control condition in which no novel name had been introduced for this object.) Presumably, typical exemplars are just too similar in appearance to other members of the familiar category for children to question their belief that the known name for the category applies to it. See Waxman and Senghas (1992) for further evidence that object similarity affects whether young children interpret two names as having mutually exclusive extensions.

Pragmatic Contrast. The Pragmatic Contrast account suggests that children expect speakers to be cooperative, which means that if a speaker wants to refer to something, he or she will use a mutually known way of making this reference, if one exists (Diesendruck & Markson, 2001; Clark, 1990; Gathercole, 1989). For example, if the two potential referents of "zav" in a scenario are a cup and an unfamiliar kind of object, the child decides that if the speaker had wanted the cup, he or she would have said "cup." Because the speaker used a different word, the child infers that he or she must want the other object.

Children do not only avoid accepting two labels as referents for a single object. In a seminal paper, Diesendruck and Markson (2001) demonstrated that children will also avoid accepting two facts to refer to one object. Children were presented with pairs of novel objects, and for each trial, the researchers provided a novel fact about one of the objects in the pair (e.g., "My sister gave me this."). Later, the children were shown the same pairs and asked to select one of the objects based on a different novel fact (e.g., "Can you give me the one from California?"). The researchers found that children disambiguated in the fact condition as frequently as they did in the more typical label condition, suggesting that the disambiguation effect is not constrained to one single domain (i.e., labels), but should instead be considered a domain-general phenomenon.

More recent research has suggested, however, that children do not have the same expectations about words as they do other types of referential expressions. Specifically, the expectation that a label will extend to all members of a category rather than be restricted to an individual appears not to apply for facts. For example, Jaswal and Hansen (2006) demonstrated that socio-pragmatic cues such as pointing and gaze direction disrupt children's tendency to disambiguate in the facts condition but not in the labels condition. In addition, Scofield and Behrend (2007) demonstrated a developmental trend, whereby four-year-old children successfully disambiguated in conditions that combined novel labels and novel facts, replicating Diesendruck and Markson's (2001) original findings, but younger children only disambiguated in the condition that involved strictly labels. While Scofield and Behrend (2007) used this evidence as support for domain-specificity, another recent study offered a more nuanced explanation. Kalashnikova, Mattock, and Monaghan (2014) tested two age groups of children (three- to five-years-old), and adults on a disambiguation paradigm involving novel labels and facts, like the one administered by Diesendruck and Markson (2001). The three-to five-year-old children disambiguated similarly in the labels and facts conditions, replicating previous findings. However, adults disambiguated significantly more frequently in the labels condition than in the facts condition, suggesting that the disambiguation strategy is used differently as age increases (Kalashnikova et al., 2014). The study provided further support for distinguishing the pragmatic contrast account, which involves an understanding of the process of communication more broadly, from the mutual exclusivity account, which is specifically concerned with the avoidance of overlapping labels.

While the Mutual Exclusivity and Pragmatic Contrast accounts differ with respect to the principle that children follow, both propose that children will reject the familiar object only if they detect a mismatch between the object's known label and the novel label (Halberda, 2003; 2006). Neither account specifies how children make this label mismatch decision, however. It is likely based on a comparison of phonological representations of the labels (Jarvis, Merriman, Barnett, Hanba, & Van Haitsma, 2004; Merriman & Marazita, 1995; Marazita & Merriman, 2004; Wall, Merriman, & Scofield, 2015). For example, suppose a child is told, “Get me the rasp,” and rasp is an unfamiliar noun. When looking for the referent, the child will reject familiar objects such as a hammer or wrench because their known names (i.e., *hammer* and *wrench*) mismatches *rasp* (see Jarvis et al., 2004, for a review of research on how children resolve cases of phonological similarity between a known and novel name).

Competitive Activation. A third set of accounts differs from the previous two in that it does not involve comparing representations of labels. Various competitive activation models (McMurray, Horst, & Samuelson, 2012; Merriman, 1999; Regier, 2005) consider the effect to be an emergent property of the excitatory and inhibitory connections among representations. These connections form as children hear various labels used to refer to various objects over time. When a novel label is presented, these connections cause a child’s representation of a novel object to receive more activation than his or her representation of the familiar object. These accounts do not involve a process in which a child decides that a representation of the novel label does not match a representation for the familiar object.

None of these accounts requires that children have a metacognitive representation of the steps involved in ultimately selecting the novel object. One might argue that the Pragmatic Contrast account comes the closest because it posits that the child recognizes that a speaker used a label other than the one they expected the speaker to use. However, a child could abide by this principle if they have simply developed a habit of rejecting an object as the referent of a referring expression (e.g., “one my uncle gave me”) whenever they retrieve a referring expression that differs from it (e.g., “one from California”). Likewise, a child might abide by the Mutual Exclusivity principle based on a habit of rejecting an object as the referent for a label whenever he or she retrieves a different label for the object. The child may not represent this decision as choosing between an object that is a novel kind over an object that is a familiar kind. Similarly, from the perspective of competitive activation models, a child who chooses the novel object because it receives more activation from the novel label than the familiar object does may not reflect on the cognitions that have led to this decision.

Development of Relevant Metacognition

Research on young children's judgment of their own lexical knowledge suggests that an ability to represent the disambiguation effect on a metacognitive level develops during the preschool years. When asked to judge whether various words and pseudowords (e.g., "zav") are ones they know, 2- and 3-year-olds often say they know the pseudowords, whereas 4-year-olds rarely make this mistake (Chaney, 1992; Merriman, Lipko, & Evey, 2008; Merriman & Schuster, 1991; Smith & Tager-Flusberg, 1982). A similar developmental trend has been found regarding children's tendency to report that they know the names for unfamiliar kinds of objects (Marazita & Merriman, 2004; Merriman & Lipko, 2008; Merriman et al., 2008; Wall, Merriman, & Scofield, 2015). Because younger preschoolers often fail to identify novel labels and novel kinds of objects as ones they do not know, it is unlikely that they would represent the disambiguation problem as one in which they must choose between an object they know and an object they do not know as the referent of a label they do not know.

Most investigations of disambiguation find that the effect is stronger for older children than for younger ones (Lewis & Frank, 2015; Merriman, Marazita, & Jarvis, 1995). The exceptions are studies in which children receive corrective feedback after every trial (Marazita & Merriman, 2004; Henning & Merriman, 2019) or the effect is put in conflict with a cue that favors the familiar object (e.g., the speaker pointing at the familiar object) (Gollek & Doherty, 2016; Grassmann & Tomasello, 2010; Scofield, Merriman & Wall, 2018).

Awareness of lexical knowledge has been found to be associated with a more robust disambiguation effect. In the disambiguation paradigm used by Merriman and colleagues, every trial involves first asking the child whether he or she knows a particular novel label and then asking for its referent (e.g., “Do you know what a zav is? Which one is a zav?) Children who are younger than 3 ½-years-old rarely respond to the first question by saying, “No.” They tend to say, “Yes,” or just not respond. In four separate studies involving either 3 ½ or 4-year-olds (Merriman & Bowman, 1989, Studies 1 and 2; Merriman & Schuster, 1991; Merriman, Marazita, & Jarvis, 1993, Experiment 2), the average correlation between children’s tendency to say, “No” (i.e., acknowledge their ignorance of a novel label) and their tendency to map the novel labels onto an unfamiliar rather than familiar objects was .47. Using the Stouffer method for combining the results of correlational studies (Rosenthal, 1991), Z = 3.77, p = .0002.

Wall et al. (2015) gave 3- and 4-year-olds the object “nameability” judgment task that was developed by Marazita and Merriman (2004). The child was asked whether they knew names for various objects that were either unfamiliar (e.g., an attachment for a water bed) or highly familiar (e.g., a sock). In a posttest, the child was shown the unfamiliar kinds again and asked what each was called. Usually the child gave no response or simply described the object. However, on occasion they produced either a correct name or an incorrect, but plausible name. These objects were excluded from the calculation of the accuracy of the child’s object nameability judgments. That is, a child’s accuracy equaled the average of the proportion of the non-excluded unfamiliar kinds that they had judged correctly (i.e., said, “No” [I do not know its name]) and the proportion of familiar kinds that they had judged correctly (i.e., said, “Yes’).

In two experiments, Wall et al. (2015) found the accuracy of these judgments to increase with age, replicating Marazita and Merriman (2004). Moreover, judgment accuracy predicted the strength of the child’s disambiguation effect in a cross-modal paradigm. In the latter, children received several trials in which they learned a name for a novel object, then examined two hidden objects with their hands. One was identical to the object that they had just learned to label and the other was unfamiliar. On some trials, they were asked to decide which one was the referent of the label that they had just learned. The children had little difficulty doing so. On the critical disambiguation trials, they were asked to decide which one was the referent of a novel label. Only the children who had made highly accurate object nameability judgments showed the disambiguation effect. Even after statistically controlling for age, the accuracy of children's object nameability judgments was significantly correlated with how frequently they chose the unfamiliar object on the disambiguation trials.

In the final experiment reported by Merriman and Bowman (1989), children were asked to justify their solutions to the standard disambiguation problem. Merriman and Schuster (1991) reported an analysis of how the children responded. While nearly all of the younger children either did not respond or pointed out some property of the unfamiliar object, many 4-year-olds pointed to the familiar object and said, "Because that one is a [familiar label, e.g., cup]." The 4-year-olds who offered this "familiar label" justification were more likely than the other 4-year-olds to have acknowledged their ignorance of the novel label before they selected its referent.

Marazita and Merriman (2004) extended this last result. They found that those 4-year-olds who offered the familiar label justification for disambiguation also tended to make more accurate lexical knowledge judgments than the other 4-year-olds. Both their judgments of word familiarity ("Do you know what a zav/table is?) and object nameability (e.g., "Do you know what this [unfamiliar/familiar object] is called?) tended to be more accurate. Among 2 ½ -year-olds, the tendency to offer the familiar label justification for disambiguation was also correlated with the accuracy of word familiarity judgments, but not with the accuracy of object nameability judgments. These results suggest that as children become aware of what they know and what they do not know, they also develop accurate metacognitions about the types of words and the types of objects that are involved in the disambiguation effect.

Two Metacognitive Disambiguation Tasks

Slocum and Merriman (2018) recently developed a metacognitive disambiguation task, in which children could only succeed if they had an awareness of their lexical knowledge. Three- and 4-year-olds first sorted objects according to their familiarity. With the help of the experimenter, they put the familiar objects into a bucket for things "I know" and the unfamiliar objects into a bucket for things "I don't know." After sorting the objects, the experimenter asked the children to recall the objects that they had put in the "I know" bucket. Most of the children could recall only one or two of these objects. The experimenter then removed whichever objects the child had recalled from the "I know" bucket and removed an equal number of unfamiliar objects from the "I don't know" bucket. Thus, the number of objects that remained in each bucket was the same and the children were not able to recall the names for any of them.

Several trials followed in which the experimenter told the child that he had seen a "[label]" in one of the buckets and then asked the child to decide which bucket it was. On every trial except one, the label was novel (i.e., pseudowords such as *blicket* and *zav*).

The only way the child could respond consistently correctly to such requests would be to note that the label was one they did not know and then reason that because it was one they did not know, the label's referent must be in the "I don't know" bucket. Because they could not recall the labels of any of the objects that remained in the "I know" bucket, it is unlikely that they would reject this bucket based on a mismatch between the novel label and the label of a specific object inside the bucket. It is also unlikely that a child would choose the correct bucket based on competitive activation because every object was hidden, and the child could not recall the names for any object in the "I know" bucket. In contrast, if the child represented the novel label as being unknown, it is likely that they would decide it applied to one of the objects in the "I don't know" bucket.

Slocum and Merriman (2018) found a significant effect of age, such that most 4-year-olds correctly selected the "I don't know" bucket on the novel label trials, and most 3-year-olds showed no preference for one bucket over the other. Moreover, after controlling for age and receptive vocabulary size, children's level of success on the buckets task was found to be predicted by their awareness of their own lexical knowledge. The latter was assessed by asking children to make yes/no judgments about whether they knew various words (Hartin, Stevenson, & Merriman, 2016). Half of the words were highly familiar and half were pseudowords. The set of words used for this task did not overlap with the ones presented in the buckets task.

These findings lend support to the view that as children develop a reflective awareness of their lexical knowledge, they tend to represent the problem of identifying the referent of a novel label in metacognitive terms. This way of representing the problem may further strengthen their tendency to select the unfamiliar object rather than the familiar object as the referent of the label.

While Slocum and Merriman (2018) provided evidence that older preschoolers can display a metacognitive disambiguation effect, they did not show that children consult metacognitive representations when solving an ordinary disambiguation problem. In the buckets task, children were prompted to use these metacognitions, as it was the only way to resolve the ambiguity. When it is possible to compare labels' phonological representations or compare objects' levels of activation, as it is in ordinary disambiguation problems, children's decisions about which object to select could well be based on these processes alone. It is possible that they might not also consider metacognitive representations of the labels or objects involved.

Henning and Merriman (2019) created a second purely metacognitive disambiguation task – called "disambiguation prediction" – to evaluate whether young children form general metacognitive representations of the elements of the disambiguation problem. In the disambiguation prediction task, children first solved four ordinary disambiguation problems (i.e., deciding whether an unfamiliar or familiar object was the referent of a novel label). The children were told that each problem was an example of how a game was played. Throughout the game, whenever they chose the unfamiliar object, they were told that they were correct and whenever they chose the familiar object, they were told that the other object was correct. After the children had played four rounds of the game, they were shown new pairs of familiar and unfamiliar objects and told that they were continuing with the game. For each of these pairs, the children were first asked to predict the object that would be "the right one." After making their prediction, the experimenter said, "Let's see," and then presented a novel label and asked them to select its referent. After the child made a selection, they received the same kind of feedback as in the first four rounds of the game.

Henning and Merriman reasoned that for children to come to believe that the correct choice would always be the unfamiliar object, they would need general metacognitive representations of their solutions to the first four problems. That is, they would need to realize that on each of those first four trials, the correct object was always one that had evoked a particular kind of cognitive experience (e.g., a feeling of novelty or an inability to retrieve a label), and the incorrect object was always one that had evoked a contrasting kind of cognitive experience (e.g., a feeling of familiarity or the retrieval of a label). They would also need to reason that if two objects that contrasted in the same way were presented on a new trial, the correct choice would be the one that evoked the kind of cognitive experience that the previous correct choices had evoked.

Henning and Merriman (2019) found a pattern of results similar to that of Slocum and Merriman (2018). While both 3- and 4-year-olds performed well on the first four ordinary disambiguation problems, only the older group made accurate disambiguation predictions. These results add further support for the hypothesis that as children get older and develop reflective awareness, they also develop an ability to represent the disambiguation problem in metacognitive terms.

The disambiguation prediction task differs from Slocum and Merriman's (2018) buckets task in several important ways. In the buckets task, the objects are hidden and the novel label is presented on each trial. In the disambiguation prediction task, the opposite is true. The objects are visible, but the novel label is not presented. In addition, passing the disambiguation prediction task does not necessarily require consciously representing the objects as ones "I know" or "I don't know." Unconscious metacognitive representations (Karmiloff-Smith, 1986) may be sufficient. For example, children may have merely noticed that the correct answer had always been the object that felt novel rather than the one that felt familiar, and then made correct predictions based on this observation.

Slocum and Merriman (2018) found performance on the buckets task to be associated with the accuracy of word knowledge judgment (e.g., “Do you know what a blicket/table is?”). In contrast, Henning and Merriman (2019) found no relation between performance on their disambiguation prediction task and word knowledge judgment. Instead, they found their task to be related to the accuracy of children’s object nameability judgment (e.g., Do you know what this [familiar/unfamiliar object] is called?). Children who made every object nameability judgment correctly also made more correct disambiguation predictions than children who made at least one erroneous object nameability judgment. It is possible that successful disambiguation prediction depends on representing the metacognitive contrast between unfamiliar and familiar objects, while successful performance on the buckets task depends on representing the metacognitive contrast between unfamiliar and familiar labels.

One limitation of Henning and Merriman’s (2019) disambiguation prediction task is that although children needed to consult general metacognitive representations to consistently make correct predictions, it is possible that they did so in the task only because they were prompted to consider what the first four problems “in the game” had in common. There is no evidence that metacognition influences any child’s processing of an ordinary disambiguation problem, that is, a problem in which a novel label is presented, the object choices are visible, and no other information about the problem is presented. Henning and Merriman (2019) did demonstrate, however, that 4-year-old children can recognize a metacognitive pattern that a set of instances follows, and then use this pattern to predict which choice will be correct on a future instance. At the very least, this conclusion, which could not be drawn from the success of 4-year-olds in Slocum and Merriman’s (2018) buckets task, strengthens the case for the view that metacognitive representations account for the increasing robustness of the disambiguation effect.

Chapter 2. Overview of the Current Investigation

The goal of the current investigation was to test the hypothesis that the development of metacognitive representations during the preschool years strengthens children's disambiguation effect. The results of the metacognitive disambiguation studies provide evidence that sometime after the fourth birthday, most children can solve novel name mapping problems that require metacognitive representations. To pass the buckets task (Slocum & Merriman, 2018), children need to be able to consistently represent novel names as names "I don't know" or as names "for something I don't know". To pass the disambiguation prediction task (Henning & Merriman, 2019), children need to represent the difference between pairs of familiar and unfamiliar objects in general metacognitive terms (e.g., as a contrast between ones that feel familiar and ones that feel unfamiliar). However, it is still an open question whether a) children ever spontaneously represent the elements of an ordinary disambiguation problem metacognitively and b) whether such representations strengthen their tendency to map the novel name to the unfamiliar object.

I addressed these questions by examining whether increasing the number of familiar choices in a disambiguation problem affects the solution processes of children who are aware of their lexical ignorance differently from children who are not aware of their lexical ignorance. Based on previous work on the disambiguation effect, I hypothesized that as the task becomes more demanding, children who are aware of their lexical ignorance will show an increasing accuracy advantage over those who lack such awareness.

My assumption is that a child who does not consult metacognitive representations will only solve disambiguation problems consistently if they consistently retrieve a label for each familiar object, decide that each label mismatches the novel label, and finally decide to select the unfamiliar object because it is the only object for which a label mismatch is not detected. Merriman and Marazita (1995) proposed a model of this kind for how 2-year-olds approach a disambiguation problem. The implication of this model is that the number of processes that the child needs to carry out correctly in order to consistently select the unfamiliar object increases as the number of familiar objects in the problem increases. Therefore, the likelihood of failing to carry out all of these processes correctly should increase, which in turn should increase the likelihood of selecting the wrong object.

Now suppose a child represents the novel name as one “I don’t know” or as one that feels unfamiliar, represents the unfamiliar object in matching terms, and represents the familiar objects in contrasting terms, that is, as ones “I know” or as ones that feel familiar. This child’s likelihood of successfully rejecting each familiar object may be greater because the child can reject them based on both mismatching cognitive representations (i.e., labels) and mismatching metacognitive representations. That is, they can reject the familiar object not only because its label mismatches the novel label, but also because it is one they know or that feels familiar, whereas the label is not one they know or is not something that feels familiar.

It is also possible that metacognitive representations will reduce the likelihood of error in the execution of the lower level cognitive processes themselves. For example, a child who does not immediately retrieve a label for a familiar object may be compelled to continue searching for it by their metacognitive representation of the object as one I know (or as one that feels familiar). Likewise, a child who tries to retrieve a label for the unfamiliar object may be compelled to stop searching for it by their metacognitive representation of the object as one I do not know (or as one that does not feel familiar).

When a disambiguation problem involves only one familiar object, label retrieval and comparison processes may be sufficient by themselves for a child to consistently decide to reject this object and select the unfamiliar object. However, as the number of familiar objects increases, these processes may no longer be sufficient by themselves to prevent an erroneous selection.

As an illustration, suppose the child's likelihood of carrying out any single cognitive process is very high, for example, .97. If the child only needs to carry out three of these processes, the chance of carrying out all three successfully is $(.97)^3 = .912$. Suppose the support from metacognitive representations increases the likelihood of carrying out any single process to .99. If only three need to be carried out, then the likelihood of carrying out all three successfully is $(,99)^3 = .970$. So there would not be much improvement. But if six of these processes need to be carried out successfully, then the likelihood of success with cognitive process alone is $(.97)^6 = .833$, whereas the likelihood of success with the support of metacognitive representations is .941. Because of these multiplicative effects, the difference in success rate between cognitive processes alone and cognitive processes with the support of metacognitive representations increases as the number of processes that need to be carried out successfully increases.

Additionally, children who consult metacognitive representations may solve these problems more quickly than children who do not. These children may interpret the presentation of a novel label as a signal to scan the set of the objects for the one that is also novel. These children may well choose correctly more quickly than children who must select referents based on a process of eliminating the familiar objects (Halberda, 2006). The metacognitively-advanced children might even show something akin to the pop-out effects that have been demonstrated in perceptual search tasks (Treisman, 1985; Wolfe, 1994). The effect occurs when a unique visual target (e.g., a red circle) is rapidly detected among a set of homogenous distractors (e.g., green circles).

Alternatively, it is possible that children who consult metacognitive representations take longer to respond. These children may be more likely to reflect on which objects they know and do not know, and they may ultimately decide more slowly than children who are not capable of engaging in this type of reflection.

Very few studies have addressed the effect of number of familiar objects on children's solution to disambiguation problems. Evey and Merriman (1998) administered a block of four problems involving only one familiar object (two choices) and a block of four problems involving three familiar objects (four choices) to 25-month-olds. The choices were actually drawings of objects rather than actual objects. The order of blocks was counterbalanced. Type of feedback was also manipulated. Some children received mild acceptance (i.e., “OK,”) for every selection they made. Others received strong acceptance (i.e., “You’re right!”) for every selection (even incorrect ones). A third group received training feedback whereby correct selections were strongly accepted and incorrect selections were corrected (“No, this one [pointing to the unfamiliar object] is the dax.). The effect of problem type depended on the type of feedback and block order.

The effect of the number of familiar objects was evident on the very first trial. For those tested on a block of two-choice problems first, nearly every child (89%) chose correctly on the first trial. In contrast, for those tested on a block of four-choice problems first, only 65% chose correctly on the first trial. Among children who received the two-choice problems first, those who received either strong acceptance or training feedback maintained a high rate of correct selection over the remaining two-choice problems, whereas those who received mild acceptance showed a modest decline (*M* = .74 correct). Among children who received the four-choice problems first, the effect of feedback was even more pronounced. Those who received training feedback increased their rate of correct selection over the remaining four-choice problems (*M* = .83). In contrast, those who received either mild or strong acceptance decreased their rate of correct selection substantially (*M* = .37). Thus, the disambiguation effect was weaker for four-choice problems than for two-choice problems on the very first trial, and this difference was magnified on subsequent trials among those who received non-contingent feedback.

Horst, Scott, and Pollard (2010) conducted the only other study of the disambiguation effect that manipulated the number of familiar objects. While this study was primarily focused on later retention of the initial name-object mappings, the first phase of their task closely resembled a typical disambiguation paradigm. In this phase, 30-month-old children were presented with a referent selection task, which included trials with three, four, or five total objects from which to choose (one novel object). Like the feedback that Evey and Merriman (1998) used in their training condition, the experimenter responded to the child's selection on a trial by pointing to the target object and naming it (e.g., "Look, this is the fode!") That is, the feedback indicated that the unfamiliar object was the correct choice, regardless of whether the child had selected it.

Overall, the number of familiar objects present did not affect the strength of the disambiguation effect, as children in all three conditions chose the target object at a higher rate than would be predicted by chance. This pattern of results is like the one exhibited by the 2-year-olds who received training feedback in Evey and Merriman's (1998) experiment. It is important to note, however, that Horst et al. (2010) did not report first trial performance, which was the only trial in the name training feedback condition of Evey and Merriman in which there was an effect of problem type. It is possible that there was an effect on these trials, which was later negated by the positive effect of training feedback on the rest of the trials. In addition, Horst et al. familiarized their participants with all of the familiar choice objects, but none of the unfamiliar choice objects in a warm-up task. This procedure may have increased their participants' disambiguation effect because the correct choice (the unfamiliar object) was the only object that had not been seen before the test trial (i.e., it was both a novel token and a novel type). Several studies have shown that 2-year-olds not only tend to map a novel name onto novel type over a familiar type, but also to map a novel name onto a novel token rather than a familiarized token (Horst, Samuelson, Kucker, & McMurray, 2011; Merriman & Bowman, 1989, Study 1; Merriman & Schuster, 1991).

Although Horst et al. (2010) found no effect of number of familiar object foils on children's disambiguation, they did find an effect of this variable on children's retention of the mappings that they had made during disambiguation. Five minutes after the final disambiguation trial, children were asked to select the referents of the labels they had mapped during the disambiguation phase. They had to select these referents from a set comprised of every novel object that had appeared as a target in the disambiguation phase. On these retention trials, children received mild acceptance feedback (i.e., "OK" or "Thank you" in response to whatever choice they made). The results showed that the children who had received three-choice disambiguation problems (i.e., ones containing only two familiar object foils) retained the mappings that they had made during disambiguation, but those who had received either four- or five-choice disambiguation problems did not. So, although children in these latter conditions had chosen a garlic press as the referent of zav, for example, there was no evidence that they remembered this mapping five minutes later. While this is an interesting finding, my primary focus is on how the number of familiar object foils affects children's disambiguation effect, that is, their tendency to correctly map the novel name in the first place. Nevertheless, my second experiment will include a retention test in addition to a test of the disambiguation effect.

Evey and Merriman (1998) and Horst et al. (2010) appear to be the only published studies to have examined the effect of number of familiar objects on children's novel name mapping. No other studies of this kind are described in a recent extensive meta-analysis of the disambiguation effect by Lewis et al. (2019). No study has examined the effect of number of familiar objects on novel name mapping in children who are older than 30 months. Thus, there are no studies of this effect on children in the age range in which metacognitive representations have been hypothesized to develop (Henning & Merriman, 2019; Slocum & Merriman, 2018).

Experiment 1 examined 3- and 4-year-old children's response time and accuracy on disambiguation trials in which novel objects appeared in arrays of varying numbers of familiar objects. Instead of physical objects, the stimuli consisted of photographs of objects displayed on a touchscreen tablet. These were positioned in an array so that the size of the pictures was not affected by how many pictures were present. In addition to the disambiguation task, children completed two tasks used to assess their awareness of their word knowledge: one in which they decided whether various objects had known names and one in which they decided whether they knew what various words and pseudowords denoted. For the reasons I have already presented, I predicted that children who demonstrated an awareness of their lexical knowledge would respond more accurately than children who lacked this kind of awareness and that the difference between these groups would increase with the number of familiar object foils. In addition, I examined whether children's speed of response to disambiguation problems depended on the number of familiar choices in the problems and whether this relationship was moderated by awareness of lexical knowledge.

Experiment 2 was like the first experiment, but with a few changes. In Experiment 1, performance was near ceiling levels on two-choice items. Additionally, results in both accuracy and response time suggested that eight-choice items may have been too difficult for these age groups. A major goal in the second experiment was to see whether the results for the middle range of number of choices would replicate. Therefore, whereas the first experiment included arrays with two, four, six, and eight choices, the second experiment included arrays with three, four, five, and six choices. The one other change was that a label mapping retention test like the one used by Horst et al. (2010) was administered after the disambiguation test.

Experiment 3 included a version of the disambiguation task that was like the one used in the second experiment. The main difference was that children's eye movements were recorded as they attempted to solve the disambiguation problems. These data may provide insight into why increasing the number of choices in a disambiguation problem tends to both reduce children's accuracy and increase the time they take to make a choice. The data may also provide insight into why these effects of number of choices tend to be moderated by individual differences in awareness of lexical knowledge.

Chapter 3. Experiment 1

The experiment was designed to test several hypotheses concerning preschool-age children's disambiguation effect. First, the strength of the effect was expected to decrease as the number of familiar objects increased (Evey & Merriman, 1998). Second, both awareness of lexical knowledge and age were expected to predict the overall strength of the effect, replicating previous studies (Merriman & Schuster, 1991; Merriman & Bowman, 1989; Marazita & Merriman, 2004). Most importantly, awareness of lexical knowledge was predicted to moderate the impact of number of familiar objects on the strength of the effect. Number of familiar objects was predicted to have less impact on children who showed greater awareness of lexical knowledge than on those who showed less such awareness. Thus, the relation between awareness of lexical knowledge and the disambiguation effect was expected to increase in strength as the number of choices increased. This moderating effect was also predicted to be independent of any moderating effect of age.

A secondary goal was to examine whether awareness of lexical knowledge or age were related to how quickly children made novel name mapping decisions. I examined relations to mean latency as well as to change in latency as number of choices increased. I also examined these same kinds of relations for familiar name mapping. Familiar name mapping serves as a useful comparison to novel name mapping. The two kinds of mapping have some processes in common (e.g., encoding of familiar kinds of objects), but also some unique processes (e.g., encoding of a familiar versus a novel label). Differences between how the latencies of these two kinds of mapping change as number of choices increase, or whether these differences depend on awareness of lexical knowledge or age, might provide insight into the processes involved in name mapping as well as in judging one's own lexical knowledge.

Lipowski and Merriman (2011) found a relation between a measure of awareness of lexical knowledge and the efficiency of processes involving familiar labels. Specifically, they found that object recognition accuracy and speed of object naming were both associated with the accuracy of object nameability judgment, but not with the accuracy of word knowledge judgment or other theory-of-mind judgments. This finding provided support for Merriman and Lipko's (2008) dual criterion account. According to this account, the efficiency of a specific memory process influences how early in development a child learns to base knowledge judgments on the outcome of that memory process. For example, children who retrieve object labels more rapidly than other children are hypothesized to be among the first to develop a robust tendency to judge an object's nameability based on whether they can retrieve a label for the object. In the current investigation, therefore, the speed of selecting the object in an array of familiar objects that corresponds to a familiar label may be correlated more strongly with the accuracy of object nameability judgment than with the accuracy of word knowledge judgment.

In Experiment 1, preschoolers completed trials in which they selected the referent of a novel label or a familiar label from an array of depicted objects. The arrays contained either two, four, six, or eight objects. Every object had a name that the children knew, except for one of the objects in the arrays for the novel label trials (i.e., the correct choice on these trials). The children also completed two measures of their awareness of lexical knowledge. In the word knowledge judgment task, they reported whether they knew what various familiar words or pseudowords denoted (e.g., "Do you know what a [e.g., table/dax] is?") (Chaney, 1992; Merriman, Lipko, & Evey, 2008; Smith & Tager-Flusberg, 1982). In the object nameability judgment task, they reported whether they knew what various familiar or unfamiliar objects were called (Merriman et al., 2008; Marazita & Merriman, 2004).

Method

Participants

Thirty-four children (20 males; *M* = 51 months, range = 38-65 months) participated. Two additional children were excluded due to failure to follow directions. All children were recruited from middle-class regions of northeast Ohio. Nearly all were Caucasian, and all were monolingual speakers of English. Each child received a few stickers for participating.

Materials

The referent selection task was completed using a Microsoft Surface 3 tablet that had a 10.8-inch touchscreen with 2 gigabytes RAM and a Quad Core Intel Atom x7-Z8700 processor. The task was programmed using OpenSesame programming software (Mathôt, Schreij, & Theeuwes, 2012). The stimuli consisted of color pictures of 120 objects. All but twelve were photographs of common objects, found by searching Google Images and selecting a picture that showed the object clearly in front of a white background. Each was judged to have a name that even the youngest children in the study would know based on norms for the typical receptive vocabularies of three-year-old children (Wooden, 2006). The names of these objects are listed in Table 22.

The twelve less common objects were chosen from objects that had been used as unfamiliar kinds in previous studies of preschool-age children (Slocum & Merriman, 2018; Henning & Merriman, 2019). These included a red massager, a steel nutcracker, a metal and plastic cleat wrench, a green sprinkler, a red tomato corer, a metal whisk, a staple-remover, a blue stethoscope, a garlic press, a green liquor pourer, a red c-clamp, and an orange garden hose connector.

Twenty-four slides of either two, four, six, or eight objects were created, six of each type. Each slide could be divided into eight regions that formed a matrix of two rows and four columns. Each object in a slide was centered in a different one of these regions. The region to which an object was assigned was randomly determined. For slides that contained fewer than eight objects, some regions were empty. (See Figure 1).

Twelve of the slides were used for familiar label tests. Every object on these slides was familiar. The labels used in these tests were *house, pencil, couch, table, guitar, hammer, book, hat, scissors, watch, jacket*, and *ball.* The remaining slides were used for novel label tests. One of the objects on each of these slides was unfamiliar and the rest were familiar. The labels used in these tests were *blicket, jegger, zav, mido, jeet, wug, cobe, pilson, borp, lide, tigg,* and *ferp*.

The materials used in the word knowledge judgment task included five familiar words (*shoe, dog, truck, chair, house*) and five pseudowords (*hust, mave, gock, prad, blim*). The object nameability judgment task included five familiar objects (a flashlight, a fork, a toy car, a key, and a sock) and five objects that were likely to have unknown labels (an egg slicer, a sanded wooden shape, a spouncer, a latch hook, and a heel cushion). These objects were used in previous administrations of the object nameability judgment task (see Hartin et al., 2016; Henning & Merriman, 2019).

Procedures

The experimenter sat opposite a child at a small table in a quiet room at the child's preschool. The child completed the word knowledge judgment, object nameability judgment, and referent selection tasks. Half of the children completed object nameability first, and half completed referent selection first. Word knowledge judgment always followed object nameability judgment. Before participating in any of the tasks, the child colored with the experimenter for 5 minutes. The session lasted approximately 20 minutes.

Referent Selection. The child was told, "We're going to play a game with this computer where you have to touch whatever object you hear. You're going to hear a word and then repeat it. Once you repeat the word, the computer will tell you to touch the picture that looks like the word you hear as fast as you can. There will be other pictures on the screen too, so you have to make sure to pick the right one." The experimenter then used the tablet to run the programmed experiment, which was automated from that point on. That is, the experimenter did not operate the tablet again until the program ended.

At the beginning of each trial, the child saw an emoticon in the center of the screen. The emoticon had a hand up to its ear, appearing to listen. The tablet played an audio recording of the experimenter saying, “The word is [target word]. What’s the word?” One second after the audio recording ended, the screen changed to show a green button in the center. The experimenter reminded the child to repeat the word and, once they did, told the child to touch the green button. This touch response prompted the program to present a blank screen with a cross in the center and play an audio recording of the experimenter saying, “Touch the [target word].” One second after the end of this recording, the screen changed to show one of the test slides. Once the child selected an object by touching the screen, the emoticon reappeared and the next trial began. The child received no feedback regarding whether they had made a correct selection. After twenty seconds, if the child had still not touched any object on the screen, the trial ended. The trial was scored as incorrect, and the next trial began. Occasionally, the experimenter encouraged the child to stay on task (e.g., “You are doing a good job.”).

Trial order was random, with three qualifications: the first and last trials were always familiar label trials, labels of the same type (i.e., familiar vs. novel) never occurred more than twice consecutively, and trials with the same number of objects in the array did not repeat until all levels were presented once. For example, once a two-object trial was presented, the next instance of a two-object trial did not occur again until a four-, six-, and eight-object trial had been presented within the same block. Half of the children were tested with one trial order, and half were tested with the reverse order.

The tablet recorded response time, which began when an array of objects was presented and ended when the child touched the screen. The tablet also recorded which object the child selected, based on the region in which the object was positioned. However, the experimenter also manually recorded which object the child selected on a separate piece of paper. No child ever touched a region with no object in it.

Object Nameability Judgment. The child was told, “Are you ready to play a game? I’m going to show you an object and you’re going to tell me if you know the name for it. If you know the name, say “Yes”. If you don’t know the name, say “No”. I’ll show you how it’s done.” The experimenter then presented a shoe and said, “Let’s see, do I know the name for this? Yes, I do. It’s a shoe. So I would say, “Yes, I know its name. I know what it’s called. Let’s try another one.” The experimenter then presented a tire pressure gauge and said, “Hmm, do I know the name for this? No, I don’t. I don’t know what this thing is called. So I would say, “No, I don’t know its name.” Now, you try it.” The experimenter then presented the test objects one at a time in a random order and asked the child, “Yes or no. Do you know the name for this?” The experimenter responded (e.g., “Very good”) regardless of the child’s response. Once the responses were recorded, the experimenter conducted a post-test for any novel object for which the child said he or she knew the name, saying “You said you know the name for this. What is it called?” In eight instances, it was evident that the child had misidentified a novel object as a kind that he or she knew (e.g., calling the spouncer a “sponge”). This trial was excluded from computation of the proportion of the child’s knowledge judgments that were correct. (See Hartin, et al., 2016, for further details about the procedures in this task).

Word Knowledge Judgment. A child was told, “I’m going to play a Yes-No game with you. I’m going to say some words. Listen carefully because some of the words are ones that you know and some are ones that you don’t know. I’m going to say a word, and then I’m going to ask you whether you know it. Just say ‘Yes’ or ‘No.’ Let me show you how to play.” The experimenter then administered two practice trials. The child was told, “The first word is *book*. Do you know what a *book* is? Say, ‘Yes.’ *Book* is a word that you know. You’ve heard that word before. You know what a *book* is. Are you ready for another one? The next word is *zimbiddy*. Do you know what a *zimbiddy* is? Say, ‘No.’ *Zimbiddy* is not a word that you know. It’s a made-up word. You’ve never heard that word before. There is no such thing as a *zimbiddy*. The word knowledge judgment task followed. The child was asked, “Do you know what a _____ is?" regarding the five familiar words and five pseudowords. Order of presentation was random, except that words of the same type never occurred more than twice in a row. The tester responded to the child’s answers with mild acceptance (e.g., “OK” or “Good”). Once the responses were recorded, the experimenter conducted a post-test for any pseudoword the child said he or she knew, saying “You said you know what a _____ is. What is a _____? Can you tell me anything about it?” In three instances, it was evident that the child had misidentified a pseudoword as a similar-sounding familiar word (e.g., “blimp” for *blim*). This trial was excluded from computation of the proportion of the child’s knowledge judgments that were correct.

Results and Discussion

Children’s mean proportion correct in each task is summarized in Table 1. Results for the familiar label trials in the referent selection task are not listed because performance was at ceiling (*M* = .99 for all problem types). In both lexical knowledge judgment tasks, children’s primary error was to claim to know unfamiliar kinds of stimuli rather than to deny knowing familiar kinds of stimuli. This error pattern is typical for these kinds of judgments (Marazita & Merriman, 2004; Henning & Merriman, 2019; Hartin et al., 2016), as well as for other kinds of knowledge judgments (Aguiar, Stoess, & Taylor, 2012; Sodian & Wimmer, 1987).

Task intercorrelations. Task intercorrelations and correlations with age (in months) are summarized in Table 2. Performance on the two lexical knowledge judgment tasks was significantly intercorrelated, even after statistically controlling for age, partial r (31) = .51, p = .002, which also fits with previous findings (Hartin et al., 2016; Lipowski & Merriman, 2011; Merriman & Lipko, 2008).

The accuracy of children’s novel label mapping was significantly correlated with both word knowledge judgment and object nameability judgment. These correlations were significant even after statistically controlling for age – for the correlation involving word knowledge judgment, partial r (31) = .38, p = .029; for the one involving object nameability judgment, partial r (31) = .54, p = .001. The latter result extends a finding that Wall et al. (2015) obtained in two separate studies. In each one, preschoolers’ object nameability judgment was positively correlated with the strength of the disambiguation effect in a cross-modal paradigm, independent of age. Regarding the former result, no study has examined whether the disambiguation effect is related to how accurately children respond when asked whether they know what various familiar and novel words denote. However, several studies have found that those 3 ½ - and 4-year-olds who tend to say, "No," when asked whether they know what some novel word denotes are more likely than their agemates to map that word onto an unfamiliar rather than a familiar object (Merriman & Bowman, 1989, Studies 1 and 2; Merriman & Schuster, 1991; Merriman et al., 1993, Experiment 2).

Age trends in accuracy. Correlations between child age and the accuracy of lexical knowledge judgments were not significant, although they were in the expected direction. The size of the correlation between age and word knowledge judgment ($r = .31$, $p = .07$) is close to the value reported in some studies of 3- and 4-year-olds (Merriman et al., 2008, Study 2; Henning & Merriman, 2019, Study 2). The average value of the correlation between age and word knowledge judgment for all comparable past assessments is .50 (This average is based on 17 samples - Chaney, 1992; Hartin et al., 2016, four samples; Henning & Merriman, 2019, two samples; Lipowski & Merriman, 2011; Marazita & Merriman, 2004; Merriman & Lipko, 2008; Merriman et al., 2008, three samples; Smith & Tager-Flusberg, 1982; Slocum & Merriman, 2018, three samples.).

The size of the correlation between age and object nameability judgment ($r = .23$, $p = .19$) is also close to the value reported in some studies of 3- and 4-year-olds (Merriman et al., 2008, Study 1; Henning & Merriman, 2019, Study 2). The average value of the correlation between age and object nameability judgment for all comparable past assessments is .37 (This average is based 12 samples - Hartin et al., 2016, four samples; Henning & Merriman, 2019; Lipowski & Merriman, 2011; Marazita & Merriman, 2004; Marazita & Merriman, 2011; Merriman & Lipko, 2008; Merriman et al., 2008; Wall et al., 2015, two samples.).

The accuracy of children's novel label mapping (i.e., the disambiguation effect) was also not significantly correlated with age. Although the trend was in the expected direction, the correlation was quite small ($r = .10$). A ceiling effect for two-choice problems ($M = .98$) does not explain this low correlation. For the correlation between age and accuracy on disambiguation problems involving four choices or more, r (32) = .12, p = .50. The average value of the correlation between age and accuracy on disambiguation problems in all comparable studies of 3- and 4-year-olds is .30 (This result is based on five samples - Beverly & Estis, 2003; Deak, Yen, & Pettit, 2001; Gollek & Doherty, 2016, Experiments 2a & 3; Merriman & Bowman, 1989, Study 1. It excludes studies that presented cues that opposed the disambiguation effect, e.g., Jarvis et al., 2004, or that provided corrective feedback during the test, e.g., Henning & Merriman, 2019.).

In sum, all three measures tended to be less strongly associated with age than in past studies. One possible explanation for the consistency of this result across the measures is that the sample just happened to consist of more advanced younger children and/or fewer less advanced older children than in past studies.

Effect of problem size on the accuracy of novel label mapping. A 4 (problem size: 2 vs. 4 vs. 6 vs. 8 choices) repeated measures ANOVA of proportion correct on the novel label trials in the referent selection task was conducted. Mauchly's test indicated that the assumption of sphericity had been violated, χ^2 (5) = 13.94, p = .016, so degrees of freedom were corrected using the Huyn-Feldt estimate of sphericity (ε = .91). The main effect of problem size was significant, F (2.37, 89.85) = 50.88, $p < .001$. As predicted, accuracy declined as the number of familiar objects in the problem increased, F (1,33) = 145.7, $p <.001$, η_p^2 = .81, for the linear contrast. The cubic contrast was also significant, F (1,33) = 5.02, p =.032, η_p^2 = .13, as the decline was sharper from the 4- to the 6-choice problems than from the 2- to the 4-choice or from the 6- to the 8-choice problems. Performance exceeded chance on every problem type, all t (33) > 6.55, $p < .001$.

My main prediction was that awareness of lexical knowledge would moderate the linear effect of problem size on accuracy of novel name mapping. To test this prediction, the slope of the linear trend over problem size in each child's performance was treated as the criterion variable in a regression analysis. The predictor variable was a composite score for awareness of lexical knowledge – the average of the child's z-score for word knowledge judgment and z-score for object nameability judgment. The regression analysis indicated that awareness of lexical knowledge did not significantly moderate the association between novel label mapping accuracy and number of total choices, $\beta = .22$, $t\ (32) = 1.30$, $p = .20$, although the relationship did trend in the predicted direction. The results were similar when the predictor variable was a single measure of awareness of lexical knowledge, for example, object nameability judgment, $\beta = .25$, $t\ (32) = 1.47$, $p = .15$, and word knowledge judgment, $\beta = .14$, $t\ (32) = .81$, $p = .43$.

In Figure 2, the change in the accuracy of novel label mapping over problem size is graphed separately for those whose awareness of lexical knowledge score was either high (at or above the median) or low (below the median). Just as predicted, accuracy declined less over the 2- , 4- , and 6-choice problems for the high-aware children than for the low-aware children. However, contrary to prediction, this trend reversed from the 6- to the 8-choice problems. In a regression analysis, awareness of lexical knowledge (the composite score) was a significant positive predictor of the slope of the trend from the 2- to the 6-choice problems, $\beta = .57$, $t\ (32) = 3.89$, $p = .001$. Those who showed greater awareness of lexical knowledge exhibited a less negative slope (less steep decline) compared to those who showed less awareness of lexical knowledge. Age was not a significant predictor of this slope, $\beta = .25$, $t\ (32) = 1.45$, $p < .156$. When the nonsignificant variance accounted for by age was removed, awareness of lexical knowledge still accounted for a significant portion of the remaining variance, R^2 change = .265, $\beta = .54$, $t\ (31) = 3.50$, $p = .001$.

In contrast, awareness of lexical knowledge was a significant negative predictor of the decline in accuracy from the 6- to the 8-choice problems, $\beta = -.38$, $t\ (32) = 2.32$, $p = .027$. This result was the opposite of what had been predicted. Those who showed greater awareness of lexical knowledge exhibited a steeper decline (a more negative slope) in accuracy from the 6- to the 8-choice problems compared to those who showed less awareness of lexical knowledge. Age was not a significant predictor of this trend, $\beta = -.10$, $t\ (32) = 0.538$, $p = .54$. When the nonsignificant variance accounted for by age was removed, awareness of lexical knowledge accounted for a significant portion of the remaining variance, R^2 change = .135, β = -.37, t (31) = 2.21, p = .034.

The inverted V pattern for the relation between awareness of lexical knowledge and problem size was somewhat more pronounced when just word knowledge judgment was entered as the predictor compared to when just object nameability judgment was entered as the predictor, although this difference was not itself significant, p > .10. For word knowledge judgment, β = .53, t (32) = 3.57, p = .001, for the trend from the 2- to the 6-choice problems, and β = -.42, t (32) = 2.63, p = .013, for the trend from the 6- to the 8-choice problems. For object nameability judgment, β = .46, t (32) = 2.95, p = .006, for the trend from the 2- to the 6-choice problems, and β = -.25, t (32) = 1.44, p = .161, for the trend from the 6- to the 8-choice problems.

Latency of referent selections on familiar label trials. A secondary goal of Experiment 1 was to examine whether the time it took children to select the referent of a label depended on problem size, child age, and child awareness of lexical ignorance, as well as on any interactions of these variables. I first describe analyses for familiar label mapping, and then for novel label mapping (i.e., disambiguation).

The distributions of the mean response latencies for each problem size were examined for outliers and normality. Outliers were defined as latencies that were more than three standard deviations different from the mean. Three were identified, each involving a different problem size and a different child. In each case, the child's response latency was unusually long. To reduce the impact of these outliers on analyses, the child's response latency was Winsorized by replacing it with a value 2.33 standard deviations above the mean for the problem size (corresponding to the 99th percentile on a normal distribution). There was no evidence that the distributions of the resulting mean latencies for any of the problem sizes deviated from normality, $|Z_{skew}|$ and $|Z_{kurtosis}| < 1.5$ for each problem size. The mean latencies are summarized in Table 3.

A 4 (problem size: 2 vs. 4 vs. 6 vs. 8 choices) repeated measures ANOVA of mean latencies on the familiar label trials was conducted. Mauchly's test indicated that the assumption of sphericity had been violated, χ^2 (5) = 26.69, $p < .001$, so degrees of freedom were corrected using the Greenhouse-Geisser estimate of sphericity ($\varepsilon = .67$). The main effect of problem size was significant, F (2.00, 66.13) = 27.50, $p < .001$. As expected, mean latency increased as the number of choices in the problem increased, F (1, 33) = 85.02, $p <.001$, $\eta_p^2 = .72$, for the linear contrast. Mean latency increased 0.33 sec on average (*SD* = 0.21) for every increment in problem size (i.e., with each addition of two choices).

Table 4 summarizes correlations between various predictors and mean latencies on familiar label trials. Age in months was a strong predictor of overall mean latency, r (32) = -.52, $p = .002$, as well as most of the problem-size specific latencies. Older children tended to take less time to map familiar names onto their referents than younger children did. Children who judged object nameability more accurately also tended to take less time to map familiar labels, compared to children who judged object nameability less accurately, r (32) = -.41, $p = .016$. There was no evidence of a relation between the accuracy of word knowledge judgment and latency to map familiar labels. When age was controlled, object nameability judgment remained significantly related to mean latency, partial r (31) = -.35, $p = .044$. The difference between this partial correlation and the one for word knowledge judgment, partial r (31) = .03, was significant, $t_{difference}$ (30) = 2.33, $p = .027$. Finally, object nameability judgment was negatively correlated with the rate at which latency increased as problem size increased (i.e., slope of the linear trend): r (32) = -.35, $p = .044$. Children who were better at judging whether various objects had known or unknown names tended to slow down less than other children did as problem size increased. However, after statistically controlling for age, this relation was not significant, r (31) = -.31, $p = .075$

The finding that object nameability judgment, but not word knowledge judgment was associated with faster responses on familiar label trials lends further support to Merriman and Lipko's (2008) proposed continuity between the efficiency of specific memory processes and the accuracy of specific knowledge judgments. According to their dual criterion account, children begin to make lexical knowledge judgments correctly when they consistently use a cue recognition and/or target generation criterion to make the judgment. For word knowledge judgment, this involves considering whether the word is one they recognize (cue recognition) or whether a meaning for the word comes to mind (target generation). For object nameability judgment, it involves considering whether the object is a kind that they recognize (cue recognition) or whether a name for the object comes to mind (target generation). The specific memory process that supports using each criterion is hypothesized to be more efficient in children who use the criterion consistently than in children who do not. In support of these claims about object nameability judgment, Lipowski and Merriman (2011) found that both the accuracy of object recognition and the speed of object naming were associated with the accuracy of object nameability judgment, but not the accuracy of word knowledge judgment or other theory-of-mind judgments. The current findings lend further support to Lipko and Merriman (2008)'s claims about the specific processes involved in the two different types of lexical knowledge judgment. How quickly a child identifies the object in a set of pictures that is an exemplar of a presented familiar label most likely depends on the efficiency of object recognition and object name retrieval processes.

Latency of referent selections on novel label trials. On a few of the novel label trials, some children did not make a choice despite having at least 20 seconds to do so. This occurred on 0%, 2%, 8%, and 10% of the 2-, 4-, 6-, and 8-choice problems, respectively. Such responses were scored as incorrect and assigned a response latency of 20 seconds.

The distributions of mean response latencies for each problem size were examined for outliers and normality. Latencies for correct and incorrect responses were analyzed separately. Outliers were defined as latencies that were more than three standard deviations different from the mean.

a) *Latency of correct responses*. Six outliers were identified, each involving a different problem size and a different child. In every case, the child took an unusually long time to respond. To reduce the impact of these outliers on analyses, the child's latency was replaced by a value 2.33 standard deviations above the mean latency for that problem size. Tests of whether the resulting distributions deviated from normality were negative, $|Z_{skew}|$ and $|Z_{kurtosis}| < 1.3$, $p > .10$ for each problem size.

The resulting mean latencies for correct responses are summarized in Table 3. Not surprisingly, children tended to take longer to map a novel label correctly than to map a familiar label correctly. This difference was significant for every problem size, all *t*s > 6.93, p < .001 (*df* varied with problem size.) Also, mean latencies to map novel labels correctly were much more variable from child to child than mean latencies to map familiar labels.

A 4 (problem size: 2 vs. 4 vs. 6 vs. 8 choices) repeated measures ANOVA of mean latencies for correct responses on the novel label trials was conducted. This analysis only involved 26 children because the other children lacked correct responses for at least one of the problem sizes. Because Mauchly's test was not significant, χ^2 (5) = 10.77, $p = .056$, sphericity was assumed. The main effect of problem size was significant, F (3, 75) = 28.58, $p < .001$. As expected, mean latency increased as the number of choices in the problem increased, F (1, 25) = 58.69, $p <.001$, $\eta_p^2 = .70$, for the linear contrast. Mean latency increased 0.89 sec on average ($SD = 0.59$) for every increment in problem size (i.e., with each addition of two choices), which is more than twice the rate for responses to familiar labels ($M = .34$), t (25) = 4.55, $p < .001$. The quadratic trend in the latencies for correct responses to novel labels was also significant, F (1, 25) = 13.05, $p = .001$, $\eta_p^2 = .34$. As can be seen in Table 3, latencies increased linearly from 2- to 4- to 6-choice problems, but then decreased slightly from 6- to 8-choice problems. The slight decrease was not significant, t (25) = 0.80, $p = .043$. The contrast between novel and familiar labels in rate of increase from 2- to 6-choice problems was striking ($M = 1.50$ and .26 sec per increment in problem size, respectively), t (25) = 5.94, $p < .001$.

The overall linear trend in the latencies mirrored the linear trend in the accuracy of novel label mapping. As problem size increased, children chose the unfamiliar object less often and took more time before selecting it. The quadratic and cubic trends did not match, however. Ceiling effects on the accuracy of 2-choice problems, but not on the latency of 2-choice problems may have accounted for some of this difference. However, whereas accuracy declined significantly from 6-choice to 8-choice problems, latencies of correct responses did not increase. One would expect children to take longer to solve problems that involve more choices. Failure to find an increase in latency from 6- to 8-choice problems is evidence that children's tendency to select an object before having checked every choice increased from 6- to 8-choice problems.

Table 5 summarizes correlations between various predictors and mean latencies of correct responses on novel label trials. Like the results for familiar label trials, age (in months) was a strong predictor of overall mean latency, *r* (32) = -.47, *p* = .002. Older children tended to take less time to identify the unfamiliar object that was the referent of the novel label, compared to younger children.

Neither word knowledge judgment nor object nameability judgment was significantly correlated with overall mean latency. However, word knowledge judgment was negatively correlated with mean latencies to solve 2- and 4-choice problems (by one-tailed tests for each problem separately and by a two-tailed test for the two problems combined, *r* (32) = -.36, two-tailed *p* = .037). Additionally, word knowledge judgment was *positively* correlated with how much the mean latencies of correct responses increased from the 4- to the 6-choice problems, *r* (28) = .55, *p* = .001. This relation remained significant after controlling for age, *r* (27) = .57, *p* = .001. Neither age nor object nameability judgment was associated with change in the latency of correct novel label responses from 4- to 6-choice problems, *r* (28) = -.11 and .19, for age and object nameability judgment, respectively, both *p* > .30.

Figure 3 displays the pattern of change for children whose scored at or above the median on word knowledge judgment (*n* = 18; *Mdn* = .90 correct) and those who scored below the median on word knowledge judgment (*n* = 16). The shift from 4- to 6-choice problems resulted in an increase of 1.18 sec in the mean latency of correct responses in the high-scoring group, but only 0.34 sec in the low-scoring group.

As already reported, word knowledge judgment was also associated with a flatter rate of decline in the accuracy of novel label responses from 2- to 4- to 6-choice problems. Children who showed greater awareness that a novel label is one that they do not know tended to slow down as much as other children as the number of choices increased from two to four. However, they tended to slow down more than other children as the number increased from four to six.

The children who showed greater awareness that a novel label is one that they do not know may not only have tended a) to carry out the processes involved in evaluating the suitability of a novel label for an individual object faster and with less error, but also b) to repeat these processes over a larger number of objects before becoming disorganized or prematurely deciding that they have identified the correct choice. Let us call a) processing efficiency and b) evaluation set limit. Greater processing efficiency would account for these children's greater speed on 2- and 4- choice problems and greater accuracy on 4-choice problems (keeping in mind that accuracy was at ceiling on 2-choice problem for nearly every child). A higher evaluation set limit could explain why their accuracy advantage becomes even greater on 6-choice problems as well as why they slow down more than the other children do. If they evaluate a larger number of the choices than the other children, then they should take longer before making their decision.

As already noted, children who scored below the median on word knowledge judgment showed only a .34 sec increase in mean latency of correct response from the 4- to the 6-choice problems. In contrast, they showed a 1.90 sec increase from the 2- to 4-choice problems, $t\ (11) = 2.01$, $p = .070$. If the evaluation set limit of most of these children was near five, for example, one would expect to find the latency increase to be greater from 2- to 4-choices than from 4- to 6- choice problems. In contrast, the children who scored above the median on word knowledge judgment showed a larger increase from 4- to 6-choice than from 2- to 4-choice problems ($M = 2.18$ and 1.18 sec, respectively), $t\ (17) = 2.08$, $p = .053$. This last finding is consistent with their evaluation set limit being six or more. It is also consistent with the possibility that as problem size approaches a child's evaluation set limit, latency of correct response shows an accelerated increase (i.e., a pronounced slowing down).

If the evaluation set limit of the below-the-median group was near five, one would expect them to show a smaller decline in accuracy from 2- to 4-choice problems than from 4- to 6-choice problems. They showed a non-significant trend consistent with this expectation. Mean proportion correct fell .21 from 2- to 4-choice problems (from .98 to .77 correct), then fell an additional .37 to .40 correct on 6-choice problems, *t* (15) = 1.33, *p* = .20. Likewise, if the evaluation set limit of the at/above-the-median group was six or more, one would expect comparable declines. Mean proportion correct fell .11 from 2- to 4-choice problems (from .98 to .87 correct), then fell an additional .22 to .65 correct on 6-choice problems, *t* (17) = 1.46, *p* = .16. Contrary to the expectation based on the hypothesized difference in evaluation set limit, the trends in the two groups were quite similar. However, interpretation of change in proportion correct is complicated by the ceiling effect on 2-choice problems, by the greater decline in the proportion correct that represents chance responding from 2 to 4 choices (from .50 to .25) than from four to six choices (from .25 to .125), and by uncertainty about the appropriate scaling of proportion correct.

Regarding the change from six to eight choices, word knowledge judgment was unrelated to the change in mean latency of correct responses on novel label trials, *r* (24) = -.30, *p* = .132. Notably, the direction of this nonsignificant relation was opposite of the significant relation found for the change from four to six choices. The decline in latency from six to eight choices was 0.49 sec (SD = 2.16) in those who scored at/above the median in the word knowledge judgment task, compared to 0.11 sec (*SD* = 1.89) in those who scored below the median, *t* (24) = .52, *p* = .61. These changes were not significant in either group.

The null correlation between word knowledge judgment and change in response latencies from 6- to 8-choice problems contrasts with the significant negative correlation found between word knowledge judgment and change in accuracy from 6- to 8-choice problems. As the novel label problems increased from six to eight choices, both mean latency of correct response and mean accuracy showed little change in the below-the-median group (for latency, M = 5.78 and 5.68 sec, respectively; for accuracy, M = .40 and .42 correct, respectively), *t*s < 1. If the evaluation set limit were near five in this group, then one would not expect a change in latency, but one might expect a modest decline in accuracy. The likelihood that they would happen to sample the novel object should be lower when there are seven competitors than when there are five competitors. It may be that when a child cannot manage to evaluate every object in a set, they remain unsure about which object is correct even if they happen to encode the novel object. In this case, they may just pick the object that they find most salient. The likelihood that the most salient object among the ones they happened to encode is the novel object may be so low that they rarely select it. Evey and Merriman (1998) found that when 2 ½-year-olds were simply asked to pick an object, they rarely picked the novel object in a set of five objects except on the very first trial.

As problem size increased from six to eight, children who scored at or above the median on word knowledge judgment showed a significant decline in mean accuracy, t (17) = 2.15, $p = .046$, and a non-significant decline in mean latency of correct responses, t (15) < 1. The decline in accuracy is consistent with the proposal that their evaluation set limit was greater than six. Regarding the change in mean latency, the proposed set limit only implies that the change once the number of choices exceeds the evaluation set limit should not be as positive as the change before the limit was reached. In this case, the change in latency from six to eight choices was not as positive as the change from four to six choices. Whether a decline, no change, or a smaller increase in latency occurs once the evaluation set limit is exceeded may depend on individual differences in reactions to not being able to evaluate every choice in a problem.

b) *Latency of incorrect responses*. Analyses of latencies for incorrect responses on novel label trials were restricted to problem sizes of 4 or more; only two children ever responded incorrectly on a two-choice problem. The distribution of the latencies for incorrect responses for the other problem types were positively skewed, Z_{skew} = 2.08, 2.17, and 2.36, for 4-, 6-, and 8-choice problem types, respectively, all $p < .05$. Because of this skew, and because standard deviations were approximately three times greater than for latencies for correct responses, mean latencies on each of these problem types were significantly shorter for correct responses than for incorrect responses, all *t*s > 6.93, $p < .001$. However, median latencies were not significantly different. For 4-choice problems, the medians were 4.88 and 4.31 sec for correct and incorrect responses, respectively, Wilcoxon signed-rank $z = 0.74$, $p = .46$. Note that this last analysis was based on only the 15 who children had at least one correct and one incorrect response, and so lacked statistical power. For 6-choice problems ($n = 24$), the medians were 5.94 and 6.10 sec for correct and incorrect responses, respectively, $z = 0.94$, $p = .35$. For 8-choice problems ($n = 26$), the medians were 5.23 and 6.30 sec for correct and incorrect responses, respectively, $z = 1.89$, $p = .058$.

Although the median latency for incorrect responses increased by 1.79 sec from 4- to 6-choice problems, this difference was not statistically significant, Wilcoxon signed-rank $z = 1.26$, $p = .21$. The statistical power of this test was weak, however; it only involved the 14 children who made at least one incorrect response on each of the problem types. The other pairs of median latencies were also not significant; for 4- vs. 8-choice ($n = 14$), $z = 0.53$, $p = .59$; for 6- vs. 8-choice ($n = 25$), $z = 0.87$, $p = .38$,

Figure 4 depicts the relation between individuals' latencies for incorrect and correct novel label responses on each problem type (Outliers were not corrected for these figures.). For four-choice problems, the latencies of the two types of responses were strongly correlated, r (13) = .59, $p = .020$. Also, 60% of the children ($n = 9$) formed a cluster that consisted of the shortest average latencies in the sample ($M = 3.70$ sec; range = 2.23 to 5.28 sec). In the remaining children, the latencies were considerably longer ($M = 11.04$; range: 7.82 to 15.83). In most of these remaining cases, incorrect responses were quite long.

A similar pattern was evident in the latencies for the six-choice problems, except that one case deviated markedly from the pattern. This child's correct response latency was more than 4 standard deviations above the mean, whereas his incorrect response latency was at the mean. When this case was removed, the correlation between latencies for correct and incorrect response was significant, r (21) = ,48, $p = .020$. Also, 67% of the children ($n = 16$) formed a cluster that consisted of the shortest average latencies (*M* = 5.30, range = 2.97 to 7.58). In the rest, the average latencies were much longer (*M* = 12.71; range: 8.80 to 15.18). In most of these remaining cases, incorrect responses were quite long.

For the eight-choice problems, one case also deviated markedly from the positive linear relation between latencies for correct and incorrect responses. This child produced the fastest correct response to an 8-choice problem in the sample (2.37 sec), but on the other 8-choice problems produced the slowest incorrect responses in the sample (both 20 sec). When this case was removed, the correlation between latencies for correct and incorrect response was significant, *r* (23) = .52, *p* = .008. Also, 65% of the children (*n* = 17) formed a cluster that consisted of the shortest average latencies (*M* = 4.67, range = 3.32 to 6.56). In the rest, the average latencies were much longer (*M* = 11.25; range: 8.90 to 14.48). In most of these remaining cases, incorrect responses were quite long.

In sum, three patterns were evident for each problem size. Children who took longer than other children to make correct selections on novel label trials also tended to take longer to make incorrect selections. Approximately 65% tended to make both responses relatively quickly, whereas the rest tended to make one or both responses relatively slowly. Typically, the average latencies of incorrect responses in the latter group were quite long. In most of these cases, the child had one or more trials in which they failed to select an object within the 20 sec time frame. In contrast, correct responses rarely took more than 10 seconds.

Chapter 4. Experiment 2

The main prediction of Experiment 1 was that awareness of lexical knowledge would be associated with a more gradual decline in disambiguation accuracy as the number of choices increased. This prediction was supported, but only when the 8-choice problems were excluded from analysis. Also, mean latency of correct disambiguation was found to increase linearly from 2- to 6-choice problems, as expected, but then decrease slightly from 6- to 8-choice problems.

Based on these results, I proposed that there was a limit on the range of problem sizes over which my main prediction would hold. For 3- and 4-year-olds, this limit is either seven or eight choices. This number represents the evaluation limit for those children in this cohort who possess the greater awareness of lexical knowledge. It is the number of choices that they can examine before becoming disorganized or prematurely deciding that they have identified the correct choice. This limit is hypothesized to be even lower for children who are less aware of their lexical knowledge.

The main goal of Experiment 2 was to further evaluate this proposal by testing children on problems sizes of 3-, 4-, 5-, and 6-choices. Because none of these problems involved more than six choices, awareness of lexical knowledge was expected to be associated with a more gradual decline in the accuracy of disambiguation as problem size increased. Moreover, by adding 3- and 5-choice problems and dropping 2-choice problems, I addressed whether my main prediction was robust over variation in the particular sizes of problems presented.

Experiment 2 had two additional goals. The first was to examine the extent to which associations with age or awareness of lexical knowledge were independent of verbal intelligence. For this reason, children were administered the Peabody Picture Vocabulary Test – Fourth Edition (PPVT-4; Dunn & Dunn, 2007), which is a test of receptive vocabulary.

The second goal was to examine children's retention of their novel label mappings. Several researchers (Horst et al., 2010; Horst & Samuelson, 2008; Bloom, 2000) have noted that while a child can successfully map a novel label onto a novel object by rejecting the competitors as potential referents, they will only retain this mapping if they encode some properties of the novel object and link these with the novel label.

In the first study to examine retention (Horst & Samuelson, 2008), 24-month-olds performed very poorly. Although they showed a strong tendency to select novel objects as the referents of novel labels, they performed no better than chance when asked after a five-minute delay to identify the particular object to which they had mapped a particular novel label. In contrast, a follow up study by Spiegel and Halberda (2011) presented 31-month-olds with referent selection phase in which children only had three seconds to initially disambiguate. They found that even with this time pressure, children not only disambiguated somewhat successfully but when later tested on whether they could pick out the referent of one of the novel labels, they did so at a rate greater than would be expected by chance.

A study by Horst et al. (2010) is particularly relevant because they found that although the number of choices presented in the initial novel name mapping test did not affect disambiguation, it did affect retention. As the number of choices increased from three to five, retention declined.

Experiment 2 included a retention task like the one used by Horst et al. (2010). The children were shown sets of three unfamiliar objects, each of which had appeared on a different novel label trial of the same problem size (either 3-choice or 4-choice). The children were tested for whether they remembered the object that had been the referent of each label.

Method

Participants

Thirty-four children (18 males; *M* = 49 months, range = 36-62 months) participated. One additional child was excluded due to failure to follow directions. All children were recruited from middle-class regions of northeast Ohio. Nearly all were Caucasian, and all were monolingual speakers of English. Each child received a few stickers for participating.

Materials and procedure

The materials and procedure were the same as those in Experiment 1, but with a few changes. First, children were administered the PPVT-4 (Dunn & Dunn, 2007) at the end of the test session. A different research assistant administered the PPVT-4 than administered the other measures. As in Experiment 1, the order of the other measures (referent selection, object nameability judgment, and word knowledge judgment) was counterbalanced. Second, the stimuli for the referent selection task consisted of the same color pictures of 120 objects, except for one of the familiar objects. The picture of an archery target that was used in Experiment 1 was replaced with a picture of a pair of trousers. Third, the 2- and 8-choice slides were replaced by 3- and 5-choice slides. As in Experiment 1, the choices were arranged such that the slide could be divided into eight regions and each object was centered in a different one of those regions. The region to which an object was assigned was once again randomly determined, and some regions were empty on every slide. The familiar and novel labels used to refer to these objects were the same as in Experiment 1.

After the child completed the familiar and novel label mapping phase of the referent selection task, the experimenter engaged the child in a short distractor task involving counting the number of fingers the experimenter was holding up. The experimenter and the child then played until five minutes had passed since the conclusion of the referent selection task. Retention of novel label mappings was then tested using the tablet. The test consisted of six trials, three for novel objects that had appeared in 3-choice arrays and three for novel objects that appeared in 4-choice arrays. On each trial, three novel objects were presented together on the screen. Object positions were randomized across trials and children were asked for the referent of a different novel label on each trial. The child received no feedback regarding whether they had made the correct selection.

Results and Discussion

Children's mean proportion correct in the tests of disambiguation, word knowledge judgment, and object nameability judgment are summarized in Table 6. As in Experiment 1, results for the familiar label trials in the referent selection task are not listed because performance was at ceiling (*M* = .98 for all problem types). In both lexical knowledge judgment tasks, children's primary error was to claim to know unfamiliar kinds of stimuli rather than to deny knowing familiar kinds of stimuli. Mean proportion correct on the test of disambiguation retention was .43 (*SD* = .28). Mean standardized score on the PPVT-4 was 116 (*SD* = 10.78).

<u>*Task intercorrelations*</u>. Task intercorrelations and correlations with age (in months) are summarized in Table 7. Two children refused to complete the vocabulary test. These children's data were not included in any correlational analyses. Accuracy scores for the two lexical knowledge judgments were significantly intercorrelated, even after statistically controlling for age and vocabulary, partial *r* (28) = .40, *p* = .028, which replicates a finding of Experiment 1 and fits with previous findings (Hartin et al., 2016; Lipowski & Merriman, 2011; Merriman & Lipko, 2008).

Disambiguation accuracy and disambiguation retention were strongly correlated, even after controlling for age and vocabulary, partial *r* (28) = .58, *p* = .001. Each of these measures was significantly associated with measures of awareness of lexical knowledge (see Table 7). Three of four of these associations remained significant after controlling for age and vocabulary: disambiguation accuracy and word knowledge judgment, partial *r* (28) = .37, *p* = .045; disambiguation accuracy and object nameability judgment, partial *r* (28) = .58, *p* = .001; disambiguation retention and word knowledge judgment, partial *r* (28) = .34, *p* = .068; disambiguation retention and object nameability judgment, partial *r* (28) = .60, *p* = .001. These findings replicate and extend the evidence from Experiment 1 for a unique association between disambiguation and awareness of lexical knowledge.

Age trends in accuracy. Consistent with most previous studies, but in contrast to the results of Experiment 1, age was significantly correlated with word knowledge judgment, object nameability judgment, and disambiguation accuracy. It was also significantly correlated with disambiguation retention (see Table 7). In Experiment 1, nonsignificant age trends in the judgments and disambiguation accuracy were in the same direction, although the one for disambiguation accuracy was quite weak ($r = .10$). The latter was significantly weaker than the one observed in the current experiment, $z = 3.02$, $p = .002$, whereas the age trends for the two judgments in Experiment 1 were not significantly different from those in the current experiment, *z*s < 1.63, *p*s < .11. The contrast in age trends for disambiguation accuracy was evident even if one just focuses on the problem types that were common to both experiments, that is, the 4- and 6-choice problems. The correlation between age and mean accuracy on these two types of problems was .09 in Experiment 1, but .70 in the current experiment.

Although sampling variability may partially explain why the age trend in disambiguation accuracy was stronger in the current experiment than in Experiment 1, older children may have benefitted more than younger ones from the elimination of the 8-choice problems in the current experiment. Children whose age was less than the median (49 months) performed similarly in Experiments 1 and 2 (*M* correct = .64 and .65, respectively), whereas those whose age was greater than or equal to the median performed worse in Experiment 1 than in Experiment 2 ($M = .73$ and .91, respectively, $t\ (33) = 3.56$, $p = .002$).

Effect of problem size on the accuracy of novel label mapping. A 4 (problem size: 3 vs. 4 vs. 5 vs. 6 choices) repeated measures ANOVA of proportion correct on the novel label trials in the referent selection task was conducted. Because Mauchly's test was not significant, $\chi^2\ (5) = 3.08$, $p = .688$, sphericity was assumed. As in Experiment 1, the main effect of problem size was significant, $F\ (3, 99) = 14.71$, $p < .001$. Additionally, as predicted, accuracy declined as the number of familiar objects in the problem increased, $F\ (1, 33) = 36.78$, $p < .001$, $\eta_p^2 = .53$, for the linear contrast. This result replicates the significant linear trend from the 2-choice to the 6-choice problems in Experiment 1. Performance exceeded chance on every problem type, all t (33) > 6.52, p < .001.

My main prediction was that awareness of lexical knowledge would moderate the linear effect of problem size on accuracy of novel name mapping. This prediction was tested the same way as in the first experiment, using a regression analysis with a composite score for awareness of lexical knowledge as the predictor variable and the linear trend as the criterion variable. Figure 5 shows the change in the accuracy of novel label mapping over problem size, graphed separately for those whose awareness of lexical knowledge was either high (at or above the median) or low (below the median). As predicted, accuracy declined less steeply for the high-aware children than for the low-aware children. The regression analysis indicated that awareness of lexical knowledge significantly moderated the linear relation between disambiguation accuracy and problem size, β = .60, t (32) = 4.21, p < .001. This finding replicates the result from Experiment 1 in the analysis that excluded 8-choice problems.

Age was also a significant moderator of the linear relation between disambiguation accuracy and problem size, β = .58, t (32) = 4.00, p < .001. (In Experiment 1, this moderation effect was not significant in the analysis that excluded the 8-choice problems, but was in the same direction, β = .25) In the current experiment, when the variance accounted for by age was removed, awareness of lexical knowledge still accounted for a significant portion of the remaining variance, R^2 change = .098, β = .39, t (31) = 2.31, p = .028. Vocabulary size was not a significant moderator of the linear relation between disambiguation accuracy and problem size, β = .20, t (32) = 1.14, p = .265.

As in Experiment 1, a secondary goal was to examine whether latency to select the referents of labels depended on problem size, child age, or child awareness of lexical knowledge, as well as on any interactions of these variables. I first describe analyses for familiar label mapping, and then for novel label mapping (i.e., disambiguation).

Latency of referent selections on familiar label trials. The distributions of the mean response latencies for each problem size were examined for outliers and normality. As in Experiment 1, outliers were defined as latencies that were more than three standard deviations different from the mean. Three were identified, each involving a different problem size and a different child. In each case, the child's response latency was unusually long. The values of these outliers were Winsorized and replaced with a value 2.33 standard deviations above the mean for the problem size (corresponding to the 99th percentile on a normal distribution). There was no evidence that the distributions of the resulting mean latencies for any of the problem sizes deviated from normality, $|Z_{skew}|$ and $|Z_{kurtosis}| < .73$ for each problem size. The mean latencies are summarized in Table 8.

A 4 (problem size: 3 vs. 4 vs. 5 vs. 6 choices) repeated measures ANOVA of mean latencies on the familiar label trials was conducted. Mauchly's test indicated that the assumption of sphericity had been violated, χ^2 (5) = 11.32, $p = .046$, so degrees of freedom were corrected using the Huynh-Feldt estimate of sphericity ($\varepsilon = .86$). The main effect of problem size was significant, F (2.59, 85.35) = 12.55, $p < .001$. Just as in Experiment 1, mean latency increased as the number of choices in the problem increased, F (1, 33) = 31.64, $p <.001$, $\eta_p^2 = .49$, for the linear contrast. Mean latency increased 0.17 sec on average ($SD = 0.18$) for every increment in problem size (i.e., with each addition of one choice). In this experiment, however, the quadratic contrast was also significant, F (1, 33) = 10.81, $p =.002$, $\eta_p^2 = .25$, as mean latency increased sharply from 3- to 4-choice problems, but leveled off as the number of choices approached 6, as can be seen in Table 8.

Table 9 summarizes correlations between various predictors and mean latencies on familiar label trials. As in Experiment 1, age was a strong predictor of overall mean latency, r (30) = -.65, $p < .001$, as well as every problem-size specific latency. Older children tended to map familiar names onto their corresponding referents more quickly than younger children did. As in Experiment 1, object nameability judgment was also related to overall mean latency, r (30) = -.36, $p = .042$, but word knowledge judgment was not, r (30) = -.19, $p = .29$. However, in contrast to Experiment 1, when age was statistically controlled, object nameability judgment was not correlated with overall mean latency, partial r (29) = .15, p = .421. Sampling error is unlikely to explain this non-replication; the latter partial correlation was statistically different from the one in Experiment 1, $z_{difference} = 1.965$, $p = .049$. It is possible that the relation is stronger when 8-choice problems are included (as in Experiment 1). When 8-choice problems were excluded, and controlling for age, the partial correlation between mean response latency and object nameability judgment in Experiment 1 was not significant, r (31) = -.26, p =.146.

Latency of referent selections on novel label trials. On a few of the novel label trials, some children did not make a choice within the 20 seconds they had to do so. This only occurred on 4% of the 6-choice problems, and never on any of the other problem types. Such responses were scored as incorrect and assigned a response latency of 20 seconds.

The distributions and mean response latencies for each problem size were examined for outliers and normality. Latencies for correct and incorrect responses were analyzed separately. Outliers were defined as latencies that were more than three standard deviations different from the mean.

a) Latency of correct responses. Four outliers were identified, each involving a different problem size and involving three total children. One child produced two of the outliers. In every case, the child was unusually slow to respond. To reduce the impact of these outliers on analyses, the child's latency was replaced by a value 2.33 standard deviations above the mean latency for that problem type. Tests of whether the resulting distributions deviated from normality were negative, $|Z_{skew}|$ and $|Z_{kurtosis}| < 1.9$, $p > .05$ for each problem size.

The resulting mean latencies for correct responses are summarized in Table 8. As in Experiment 1 and not surprisingly, children tended to take longer to map a novel label correctly than to map a familiar label correctly. This difference was significant for every problem size, all ts > 5.44, $p < .001$ (df varied with problem size). Just like in the first experiment, mean latencies to map novel labels correctly were much more variable from child to child than mean latencies to map familiar labels.

A 4 (problem size: 3 vs. 4 vs. 5 vs. 6 choices) repeated measures ANOVA of mean latencies for correct responses on the novel label trials was conducted. This analysis only included 27 children because the other children lacked correct responses for at least one of the problem sizes. Because Mauchly's test was not significant, χ^2 (5) = 10.30, *p* = .067, sphericity was assumed. The main effect of problem size was significant, *F* (3, 78) = 21.66, *p* < .001. As expected and as in Experiment 1, mean latency increased as the number of choices in the problem increased, *F* (1, 26) = 46.70, *p* <.001, η_p^2 = .64, for the linear contrast. Mean latency increased 0.61 sec on average (*SD* = 0.46) for every increment in problem size (i.e., with each addition of one choice), which is a significantly greater rate than responses to familiar labels (*M* = .17 sec), *t* (26) = 5.09, *p* < .001. As can be seen in Table 8, latencies increased linearly from 3- to 6-choice problems.

The overall linear trend in the latencies mirrored the linear trend in the accuracy of novel label mapping. As problem size increased, children chose the unfamiliar object less frequently and took more time before selecting it. This result replicates the linear trends in accuracy and latency observed in Experiment 1 for problems involving six or fewer choices.

Table 10 summarizes correlations between various predictors and mean latencies of correct responses on novel label trials. Unlike the results for familiar label trials and contrary to Experiment 1, age (in months) was not a predictor of overall mean latency, *r* (23) = -.38, *p* = .062. This result was in the same direction as in Experiment 1 (where r = -.47), however, and was not significantly different from it. This particular analysis had fairly low statistical power in the current experiment (*df* = 23); the estimated effect size (r = -.38) is in the medium-to-large range (Field, 2019).

In the current experiment, word knowledge judgment, object nameability judgment, and vocabulary size were significantly correlated with mean latency of correct responses on novel label trials. Children who made the judgments more accurately than other children also tended to map novel labels to novel objects more quickly. The same was true of children who had larger vocabularies, compared to children who had smaller vocabularies. The correlations between knowledge judgment and mean latency remained significant even after controlling for age and vocabulary size (for word knowledge judgment, partial r (21) = -.41, p = .05; for object nameability judgment, partial r (21) = -.61, p = .002.) Likewise, the correlation between vocabulary size and mean latency remained significant after controlling for age and the two knowledge judgments, partial r (20) = -.49, p = .021.).

Object nameability judgment, but not word knowledge judgment, was negatively correlated with the linear increase in mean latency of correct responses over problem size, r (23) = -.45, p = .024. Figure 6 displays the pattern of change for children who scored above the median on object nameability judgment and those who scored at or below the median on object nameability judgment. Age was not associated with the linear trend, r (25) = -.18, p = .595, but vocabulary size was, r (25) = -.49, p = .024. Object nameability judgment remained significantly correlated with the linear trend even after age and vocabulary were statistically controlled, r (21) = -.48, p = .020. The children who made more accurate judgments of object nameability showed a smaller increase in response latency as problem size increased, compared to those who made less accurate judgments of object nameability.

Vocabulary size remained significantly correlated with the linear increase in mean latency of correct response after age and the two knowledge judgments were statistically controlled, r (20) = -.51, p = .014. The children who had larger vocabularies showed a smaller increase in response latency as problem size increased, compared to children with smaller vocabularies.

b) *Latency of incorrect responses*. Analyses of latencies for incorrect responses on novel label trials were restricted to problem sizes of 4 or more. Only six children ever responded incorrectly on a 3-choice problem. In contrast, 13, 19, and 21 responded incorrectly at least once on 4-, 5-, and 6-choice problems, respectively. Unlike in Experiment 1, the distribution of latencies for incorrect responses for on novel label trials did not deviate from normality, $|Z_{skew}|$ and $|Z_{kurtosis}| < 1.30$, $p > .10$ for each problem size. As in Experiment 1, mean latencies were significantly shorter for correct responses (M = 5.36 sec, SD = 1.78) than incorrect responses (M = 7.85 sec, SD = 3.75) on 6-choice problems, t (15) = 3.77, p = .002. In contrast, however, mean latencies for correct responses (M = 4.00 sec, SD = 1.16) and incorrect responses

(*M* = 4.29 sec, *SD* = 1.06) did not differ on 4-choice problems, *t* (11) = 1.01, *p* = .33, and mean latencies for correct responses (*M* = 4.73 sec, *SD* = 1.85) and incorrect responses (*M* = 4.96 sec, *SD* = 1.97) did not differ on 5-choice problems, *t* (15) = 0.87, *p* = .40. Because only 12 to 17 children had at least one correct and one incorrect response on a problem of the same size, these comparisons lacked statistical power. The finding in the current experiment that mean latency of incorrect responses did not differ from mean latency of correct responses for 4-choice problems is comparable to the pattern found in Experiment 1 with respect to median latency. Likewise, the significantly shorter mean latencies for correct responses than for incorrect responses on 6-choice problems in the current experiment matches the direction of the non-significant trend observed for the median latencies on these problem types in Experiment 1.

The mean latency for incorrect responses increased by 1.14 sec from 4- to 5-choice problems and by 1.89 sec from 5- to 6-choice problems, and both of these differences were statistically significant, *t* (10) = 2.24, *p* = .049 and *t* (14) = 2.14, *p* = .05, respectively. The difference in mean latency between 4- and 6-choice problems was also significant, *t* (12) = 3.73, *p* = .003. The increase in mean latency for incorrect responses from 4- to 6-choice problems in the current experiment (3.65 sec) was greater than the increase observed in Experiment 1 for these problem types (1.20 sec), *t* (45) = 3.05, *p* = .004.

Figure 7 depicts the relation between individuals' latencies for incorrect and correct novel label responses on each problem type. (Outliers were not corrected for these figures.). For four-choice problems, the latencies of the two types of responses were strongly correlated, *r* (10) = .62, *p* = .033. A similar pattern was observed in the latencies for the five-choice problems, *r* (14) = .85, *p* < .001, and the six-choice problems, *r* (14) = .77, *p* = .001. Thus, as in Experiment 1, children who tended to take longer to make correct selections on novel label trials also tended to take longer to make incorrect selections. The main difference from Experiment 1 was that the slowest responders did not show as great a difference in the latency of their incorrect responses compared to their correct responses (compare the 4- and 6-choice graphs in Figure 4 to those in Figure 7). Consequently, the more extreme scores were closer to the other scores in the scatterplot in the current experiment than in Experiment 1.

Effect of problem size on retention of novel label mappings. As already reported, children's retention of the correct solutions to the 3- and 4-choice disambiguation problems was especially strongly related to their overall disambiguation accuracy and their awareness of lexical knowledge (see Table 7). When shown three novel objects that had been correct choices on different disambiguation trials (i.e., correct mappings for different novel labels), those who tended to remember which object had been the referent of the different novel labels tended to be those who had shown the strongest disambiguation effect and the greatest awareness of lexical knowledge.

One concern was that the relation between awareness of lexical knowledge and retention of novel label mappings was simply an artifact of their shared relation to the strength of the disambiguation effect (i.e., to how often the child mapped the novel labels correctly in the first place.) When a child made the mistake of selecting a familiar object as the referent of the novel label, they would later have to guess on the retention test for this label. However, awareness of lexical knowledge remained significantly correlated with retention of novel label mappings even after controlling for how often children had mapped these labels correctly, partial r (31) = .50, p = .003. Also, when only children who had selected correctly on every 3-choice problem were considered, awareness of lexical knowledge was still a significant predictor of retention of the novel label mappings from these problems, r (26) = .64, p < .001.

Figure 8 shows the proportion of correct selections on the retention trials graphed separately for those whose awareness of lexical knowledge score was either high (above the median) or low (at or below the median). Overall, high-aware children retained a greater proportion of the novel label mappings than low-aware children, t (32) = 4.79, p < .001. In addition, high-aware children retained more such mappings than would be expected by chance (.33) for both 3-choice problems, t (16) = 4.97, p < .001, and 4-choice problems, t (16) = 2.31, p = .034. In contrast, low-aware children failed to retain previously mapped labels that appeared in either problem type at greater than chance levels.

Horst et al. (2010) found that 30-month-olds retained a significantly higher proportion of the correct novel label mappings from 3-choice problems than from either 4- or 5-choice problems. Also, their retention only exceeded chance levels for the novel label mappings from the 3-choice problems. The authors posited that retention declined as the number of choices increased because the time spent attending to the correct choice (the novel object) declined relative to the time spent attending to incorrect choices (the familiar objects)

In the current experiment, children's retention also exceeded chance levels for the novel label mappings from 3-choice problems, t (33) = 2.39, p = .023, but not 4-choice problems, t (33) = 0.73, p = .473. However, the effect of problem size on their retention was not significant, t (33) = 1.73, p = .094. Among the high-aware children only, there was a trend toward better retention of the novel label mappings from 3-choice problems (*M* = .71, *SD* = .31) than of those from 4-choice problems (*M* = .51, *SD* = .31), but the trend was not significant, t (16) = 1.83, p = .086.

I have found awareness of lexical knowledge to be related to various aspects of the disambiguation effect. Those with higher levels of awareness show a shallower rate of decline in both the accuracy and the speed of responses as problem size increases. They also show greater retention of these novel label mappings.

Two explanations for these relations are plausible. These explanations are not incompatible. So both may partially account for the relations. The first explanation is that awareness of lexical knowledge has a direct causal impact on the disambiguation effect. Children who are aware of whether various words are ones they know and whether various objects have known labels may be more likely to note when mapping a novel label that only one of the choices matches the label metacognitively (e.g., that both are ones they do not know). This observation may cause them to select this object more often and more quickly (and perhaps even more confidently), which in turn may cause them to establish a more enduring representation of the object as an exemplar of the category denoted by the label. Noting the metacognitive match between the novel object and the novel label is not an absolute requirement for producing or retaining the disambiguation effect, but it may boost the probability of producing and retaining the effect.

This explanation is supported by previous work showing that children with greater awareness of their lexical knowledge are more likely to (a) choose from a bucket of objects whose names they do not know as the referent of a novel label (Slocum & Merriman, 2018) and (b) form general metacognitive representations based on previous instances of a “game” to predict that the correct choice is always the novel object (Henning & Merriman, 2019).

The second explanation is that awareness of lexical knowledge is related to the disambiguation effect indirectly. Some children execute basic (non-metacognitive) processes such as recognizing a familiar kind of object and retrieving its name more efficiently than other children do. Consequently, these children produce a more robust disambiguation effect because they reject every familiar object choice in a disambiguation problem more reliably and rapidly than other children do. These children may also learn at an earlier age than other children to judge their lexical knowledge accurately by reflecting on whether objects or words evoke such recognition and retrieval processes. This explanation is supported by evidence that the accuracy of object recognition and the speed of object naming are associated with the accuracy of object nameability judgment (Lipowski & Merriman, 2011; Merriman & Lipko, 2008).

Chapter 5. Experiment 3

Experiments 1 and 2 demonstrated that awareness of lexical knowledge was associated with a more gradual decline in disambiguation accuracy as the number of potential referents increased within the range from two to six. In addition, Experiment 2 showed that object nameability judgment was associated with a more gradual increase in latency of correct responses as the number of potential referents increased. One goal of Experiment 3 was to determine whether the main findings of Experiment 2 replicate.

A more important goal of Experiment 3 was to shed new light on the potential difference in strategy for solving disambiguation problems between children with awareness of their lexical knowledge and children who lack such an awareness. In Experiment 3, instead of completing the referent selection task on a touch screen tablet, children completed it on a laptop computer while a screen-based eye tracker recorded their eye movements.

Experiments 1 and 2 provided indirect evidence that awareness of lexical knowledge predicts disambiguation efficiency, especially when the task becomes more demanding. Eye tracking may help clarify the differences in children's strategies when solving these problems. For example, children with awareness of their lexical knowledge may choose the novel object as the referent for the novel label as soon as they fixate on it once, whereas children who lack this type of awareness must fixate on every object first. Additionally, there may be a particular difference in how long children spend looking at the target object, relative to the time spent looking at the familiar object foils. If children with awareness of their lexical knowledge tend to spend less relative time looking at the target object than children who lack such an awareness, it may suggest greater confidence in the selection. Finally, examining children's looking patterns qualitatively may provide a general insight into the strategy they employ when disambiguating in the presence of several familiar object foils.

Method

Participants

Forty 3- and 4-year-olds were recruited from preschools in middle-class regions of Northeast Ohio. None of the children had participated in any of the previous experiments, and all children were monolingual speakers of English. Each child received a few stickers for participating.

Materials

The referent selection task was completed on a Dell Latitude E5570 laptop computer, and each child's eye movements was recorded using a Tobii Pro X3-120 screen-based eye tracker. This eye tracker utilizes a gaze sampling frequency of 120 Hz, which indicates that it is designed for detailed research into both the timing and duration of fixations. The eye tracker fastens to the bottom of the laptop computer screen and allows for children to move their heads relatively freely. It has a gaze recovery time of less than 100 milliseconds to accommodate instances in which children momentarily look off-screen.

The referent selection task was programmed and run using Tobii Pro Studio software, which allowed for the full-screen presentation of trials as well as the recording and storage of eye movements and fixation durations. The stimuli consisted of the same arrays of color pictures as in Experiment 2. The order of trials in this task was randomized and counterbalanced in the same fashion as Experiment 2. The materials used for the word knowledge judgment, object nameability judgment, and receptive vocabulary tasks were also the same as in the previous experiments.

Procedures

The experimenter sat next to the child at a table in a quiet room at the child's preschool. The child completed the referent selection task, word knowledge and object nameability judgment tasks, and the

PPVT-4 receptive vocabulary assessment. As in the previous experiment, half of the children completed referent selection first and half completed object nameability judgment first. Once again, word knowledge judgment always followed immediately after object nameability. The PPVT-4 was always presented last. Each child completed the first three tasks with one experimenter, then moved to a new station to complete the PPVT-4 with a second experimenter. The entire session lasted between 20 and 25 minutes.

Referent Selection. This was the only task for which the procedures were changed from Experiment 2 to 3. The experimenter told the child that they were going to play a game with a computer. To calibrate the eye tracker, the child needed to follow a dot that moved to various locations on the screen. The experimenter told the child, "You're going to see a red dot on the screen. I want you to look at the dot wherever it goes. If the red dot moves around the screen, I want you to keep looking at it until I tell you to stop." If the child failed to follow the dot, the experimenter repeated the instructions and started a new calibration trial. Once the calibration phase ended, Tobii Pro Studio displayed a readout of calibration accuracy that identified any instances in which the child's eye tracking had been interrupted. When these instances occurred, the experimenter attempted to recalibrate the eye tracker for the child.

Once the child had completed the calibration phase with full accuracy, the experimenter told the child, "We're going to play a game with this computer where you have to find different things on the screen. The computer will tell you what to find, and I want you to tell me when you find it. You can point to it after you tell me that you found it, okay? Let's do a few practice tries." After each trial, the experimenter used the computer's keyboard to advance the program to the next trial. At the beginning of each trial, a cross appeared in the center of the screen. Two-hundred milliseconds after it appeared, an audio recording of the experimenter's voice played over the computer's speakers saying, "Tell me when you've found the [target word]." One second after the audio recording ended, the array of objects for the trial appeared.

After the child reported having found the object, the experimenter said "OK" and told the child to point to it. In some instances, a child spoke too softly or pointed without verbalizing. Whenever these occurred the experimenter reminded the child to first say aloud that they had found the object and then point. Three practice trials were administered to ensure that the child followed this instruction. Children were asked to wait to point because of concern that pointing might interrupt the recording of their eye gaze. The computer's microphone recorded their verbalizations, making it possible to measure the time that elapsed between the onset of the array of objects and the child's report of having found the target object. On each trial, after the child pointed to the object, the experimenter said "OK" and started the next trial. To encourage the child to stay on task, the experimenter occasionally told them, "You are doing a great job!"

Results and Discussion

Children's mean proportion correct in the tests of disambiguation, word knowledge judgment, and object nameability judgment are summarized in Table 11. Results for the familiar label trials in the referent selection task are not listed because performance was at ceiling ($M > .96$ for all problem types). In both lexical knowledge judgment tasks, children's primary error was to claim to know unfamiliar kinds of stimuli rather than to deny knowing familiar kinds of stimuli. Mean standardized score on the PPVT-4 was 110 ($SD = 9.79$), which did not differ from that of Experiment 2, $t\ (70) < 1$.

Task intercorrelations   Task intercorrelations and correlations with age (in months) are summarized in Table 12. Two children refused to complete the vocabulary test. These children's data were not included in any correlational analyses. Performance on the two lexical knowledge judgment tasks was significantly intercorrelated, even after statistically controlling for age and vocabulary, partial r (34) = .49, $p = .002$, which replicates what was observed in Experiments 1 and 2 and fits with previous findings (Hartin et al., 2016; Lipowski & Merriman, 2011; Merriman & Lipko, 2008).

As in the previous experiment, disambiguation accuracy was positively correlated with each measure of awareness of lexical knowledge (see Table 12). Unlike in the previous experiment, neither of these associations remained significant after controlling for age and vocabulary, both partial rs < .15.

Also unlike in Experiment 2, receptive vocabulary was significantly correlated with disambiguation accuracy in the current experiment. However, this association did not remain significant after controlling for age and awareness of lexical knowledge, either, partial r (34) = .14, p = .41.

Age trends in accuracy. Consistent with Experiment 2, as well as with most previous studies, age was significantly correlated with word knowledge judgment, object nameability judgment, and disambiguation accuracy (see Table 12). In the current experiment, the association between age and disambiguation accuracy was significant even after controlling for the other predictors (i.e., receptive vocabulary and awareness of lexical knowledge), partial r (34) = .46, p = .005. As I proposed regarding why this relation was significant in Experiment 2, but not Experiment 1, older children may benefit more than younger ones from the elimination of 8-choice problems. Children whose age was less than the median (49 months) performed similarly on disambiguation problems in Experiments 1, 2, and 3 (M correct = .64, .65, and .57, respectively), whereas those whose age was greater than or equal to the median performed worse in Experiment 1 (M correct = .73) than in Experiment 2 (M = .91) and Experiment 3 (M = .83), both ts > 2.0, ps < .05.

In the current experiment, children were significantly less accurate on three-choice novel label mapping problems than in Experiment 2, t (72) = 2.65, p = .01. Overall accuracy was not significantly different in the two experiments, however (M = .70, SD = .24 in Experiment 3; M = .76, SD = .23 in Experiment 2), t (72) = 1.27, p = .21. Note that on each trial in the first two experiments, the child first heard the target label, then was told to repeat the label, and then saw the array of choice objects. In the current experiment, however, the child was simply told, “Tell me when you’ve found the [target word],” and one second later, saw the array of choice objects. The greater opportunity to process the novel label (i.e., comprehend it and then repeat it back) before being asked to find its referent may have promoted more accurate referent selections on three-choice problems. It is not clear why this effect would be limited to the three-choice problems, however.

In Experiment 3, there was no evidence that this procedural change affected older children any differently than younger ones. For the children whose age was greater than the median, mean accuracy on novel label trials was .90 (*SD* =.13) in Experiment 2 and .83 (*SD* =.18) in Experiment 3. For those whose age was less than the median, mean accuracy was .62 (*SD* = .22) in Experiment 2 and .56 (*SD* = .22) in Experiment 3. Neither of these differences was significant, *ts* < 1.33*, p* > .20.

In contrast, there was evidence that eliminating the opportunity to hear and then repeat back the novel label before selecting its referent only affected the more metacognitively advanced children. For children who scored above the median on the composite measure of awareness of lexical knowledge, mean accuracy on novel label problems was higher in Experiment 2 (*M* = .91, *SD* = .12) than in Experiment 3 (*M* = .78, *SD* = .20), *t* (35) = 2.34, *p* = .025. For those who scored at or below the median on awareness of lexical knowledge, mean accuracy on novel label problems was comparable in Experiment 2 (*M* = .61, *SD* = .21) and Experiment 3 (*M* = .64, *SD* = .23), *t* (34) = .41, *p* = .69. This interaction was reflected in the significant reduction in the correlation between awareness of lexical knowledge and mean accuracy on novel label problems from Experiment 2 (*r* = .78) to Experiment 2 (*r* = .44), $z_{\text{difference}}$ = 9.28, p < .001. Thus, there was evidence that the only children who benefitted from hearing and repeating back the novel label before being asked to find its referent (the procedure unique to Experiment 2) were those who were more metacognitively advanced. Further implications of this finding are explored in the General Discussion; nevertheless, a future study should test this hypothesis by employing one condition, in which children only hear the novel label once (like in Experiment 3), and another condition, in which children hear the novel label and are given an opportunity to repeat the label before selecting its referent (like in Experiment 2).

Effect of problem size on the accuracy of novel label mapping. A 4 (problem size: 3 vs. 4 vs. 5 vs. 6 choices) repeated measures ANOVA of proportion correct on the novel label trials in the referent selection task was conducted. Because Mauchly's test was not significant, χ^2 (5) = 5.54, *p* = .353, sphericity was assumed. As in Experiments 1 and 2, the main effect of problem size was significant, *F* (3,

117) = 11.31, $p < .001$. Additionally, as predicted, accuracy declined as the number of familiar objects in the problem increased, $F\ (1, 39) = 30.97$, $p <.001$, $\eta_p^2 = .44$, for the linear contrast. This result replicates the significant linear trend observed for the corresponding trial types in both of the previous experiments. Performance exceeded chance on every problem type, all $t\ (39) > 5.99$, $p < .001$.

Once again, my main prediction was that awareness of lexical knowledge would moderate the linear effect of problem size on accuracy of novel name mapping. This prediction was tested the same way as in the prior two experiments, using a regression analysis with a composite score for awareness of lexical knowledge as the predictor variable and the slope of the linear trend as the criterion variable. Figure 9 shows the change in the accuracy of novel label mapping over problem size, graphed separately for those whose awareness of lexical knowledge was either high (at or above the median) or low (below the median). Unlike in Experiment 2 and contrary to my prediction, awareness of lexical knowledge did not significantly moderate the linear relation between disambiguation accuracy and problem size, $\beta = .14$, $t\ (36) = .88$, $p = .39$. Additionally, neither measure of awareness of lexical knowledge was found to significantly moderate this association independently. For word knowledge judgment, $\beta = .17$, $t\ (36) = 1.06$, $p = .30$. For object nameability judgment, $\beta = .06$, $t\ (36) = .63$, $p = .71$.

These findings were rather surprising. The stimuli were identical in Experiments 2 and 3. The procedures of these experiments were also very similar. If, as discussed previously, the metacognitively advanced children are the only ones who benefit from hearing and repeating a novel label before finding its referent, a procedure that eliminates this step may cause the pattern of performance by high-aware children to look more like that of low-aware children. Interestingly, age was still a significant moderator of the linear relation between disambiguation accuracy and problem size, $\beta = .36$, $t\ (36) = 2.28$, $p = .029$, in the current experiment (See Figure 10). Age also moderated this relationship in Experiment 2. Thus, although the opportunity to hear and repeat a novel label before selecting its referent may benefit children who are more metacognitively advanced, there is no evidence that it benefits children who are older.

As in the previous experiments, another goal was to examine whether latency to select the referents of labels depended on problem size, child age, or child awareness of lexical knowledge, as well as on any interactions of these variables. Once again, I first describe analyses for familiar label mapping, and then for novel label mapping (i.e., disambiguation).

Latency of referent selections on familiar label trials. The distributions of the mean response latencies for each problem size were examined for outliers and normality. As in the first two experiments, outliers were defined as latencies that were more than three standard deviations different from the mean. Four were identified, each involving a different child. In each case, the child's response latency was unusually long. The values of these outliers were Winsorized and replaced with a value 2.33 standard deviations above the mean for the problem size (corresponding to the 99th percentile on a normal distribution). In contrast to the results of the previous experiments, significant positive skew was evident in the distributions of the resulting mean latencies for each problem size, Z_{skew} ranged in value from 2.13 to 2.89, all $p < .05$. (None showed significant kurtosis, $|Z_{kurtosis}| < 1.22$ for every problem size.) The mean latencies are summarized in Table 13. To correct for the skew in these mean latencies, a log transformation was applied to them. The distributions of the transformed latencies did not show significant deviation from normality, $Z_{\text{skew}} < 1.40$ for every problem size. I report analyses of the transformed latencies, except when comparing results with the previous experiments. For the latter, I report analyses of untransformed mean latencies.

Mean latencies were faster in Experiment 3 than in Experiment 2 for every trial type, all $t\,(72) > 4.13$, $p < .001$. In Experiment 2, latency was recorded by the touchscreen tablet – the time started when the array of objects appeared on the screen and stopped when the child touched the screen to select one. Because the computer used in conjunction with the eye tracker in Experiment 3 did not have touchscreen capability, latency was recorded differently. In this experiment, children were told to tell the experimenter when they had found the object, while the computer stored an audio record of the task. Latencies were later calculated using audio-editing software – once again the time started when the array of objects appeared on the screen, but the time stopped at the first instant that the child began to say they had found the object. Because the procedure and stimuli were otherwise very similar to the previous experiment, it is likely that faster latencies observed in the current experiment resulted from the time saved by not having to reach out and touch the screen.

To explore the effect of problem size on latency, a 4 (problem size: 3 vs. 4 vs. 5 vs. 6 choices) repeated measures ANOVA of the log-transformed mean latencies on the familiar label trials was conducted. Mauchly's test indicated that the assumption of sphericity had been violated, χ^2 (5) = 13.56, *p* = .019. Therefore, degrees of freedom were adjusted by the Huynh-Feldt correction (epsilon = .861). Although the main effect of problem size did not exceed the criterion for statistical significance, *F* (2.58, 95.6) = 2.75, *p* = .055, the linear trend component was significant, F (1, 37) = 5.14, *p* =.029, η_p^2 = .12. As can be seen in Table 13, mean latency tended to increase as problem size increased, but the increase was only 0.09 sec for each additional object choice. In Experiment 2, the rate of increase was about twice as great (0.17) and the effect size was about four times greater (η_p^2 = .49). This difference may be due to measuring response latency by the time it took children to say that they had found the target object rather than by the time it took them to point to the target object. As the number of choice objects increases, the child's point necessarily needs to be more precise spatially. The more precise the point, the more time it may take for children to execute it. Eliminating the requirement to point eliminated the contribution of this motor component to response time.

Table 14 summarizes correlations between various predictors and mean latencies on familiar label trials. Unlike the previous experiments, age was not a significant predictor of overall mean latency, *r* (36) = -.19, *p* = .244, or for any other problem-size specific latency at the two-tailed level. Although sampling error may have contributed to this difference, the correlation between age and overall mean latency was significantly weaker than the correlation between these variables in Experiment 2 (*r* = -.65), $z_{\text{difference}}$ = 2.32, *p* = .020. Thus, it is likely that the change in procedure was also part of the reason why the age-related speed advantage that was evident in Experiment 2 was not evident in Experiment 3.

Two aspects of this procedural change may have been factors. One was the change from a pointing response to a verbal response (i.e., "I found it.") The age-related variance in speed to point to an object may be greater than that in speed to say "I found it." Venetsanou et al. (2009) reported that performance on the Response Speed subtest of the Bruininks-Oseretsky test of motor proficiency increased over the preschool years. The age range in their study (48 to 71 months) only overlapped partially with the range in the current investigation. However, the difference between the youngest children in their study and those a year older in performance on this test was large, Cohen's d = 1.10 (equivalent to $r^2 = .23$.)

Another potential factor is that in Experiment 2, but not Experiment 3 children had the opportunity to hear and repeat back the familiar label before being asked to find its referent. This procedure may have increased the speed with which older children identified the exemplars of familiar labels more than it increased the speed with which younger children did so.

There was no evidence that the change in procedure from Experiment 2 to 3 affected the variance in mean response latency on familiar label trials that was associated with awareness of lexical knowledge or vocabulary size. None of the correlation coefficients for these relations in Experiment 2 (see Table 9) were significantly different from those in Experiment 3 (see Table 14), max $z_{difference} = 0.69$, $p = .49$. Only the age-related variance in response latency on familiar label trials was significantly different in the two experiments.

The correlation between mean response latency of familiar label trials and the composite measure of awareness of lexical knowledge (i.e., mean of the z scores for word knowledge judgment and object nameability judgment) was fairly consistent over the three experiments: $r = -.31$, -.30, and -.32 in Experiments 1, 2, and 3, respectively. This cumulative result is statistically significant, $z = 3.15$, $p = .001$. The correlation with word knowledge judgment alone was statistically significant in Experiment 3 ($r = -.35$), but not in Experiment 1 ($r = -.14$) or 2 ($r = -.19$). However, the result for Experiment 3 was not significantly different from combined result for Experiments 1 and 2, $z_{difference} = 0.93$, $p = .35$. The correlation with object nameability judgment alone was statistically significant in Experiments 1 (r = -.41) and 2 (r = -.36), but not in Experiment 3 (r = -.22). However, the result for Experiment 3 was not significantly different from the combined result for Experiments 1 and 2, $z_{difference}$ = -0.93, p = .35. Finally, the change in the value of r from Experiments 1 and 2 to Experiment 3 for word knowledge judgment was not significantly different from the change for object nameability difference, $z_{difference}$ = 1.32, p = .19.

Latency of referent selections on novel label trials. The distributions and mean response latencies for each problem size were examined for outliers and normality. Latencies for correct and incorrect responses were analyzed separately. Outliers were defined as latencies that were more than three standard deviations different from the mean.

a) Latency of correct responses. Two outliers were identified, each was associated with a different problem size and different child. In each case, the child was unusually slow to respond. To reduce the impact of these outliers on analyses, the child's latency was replaced by a value 2.33 standard deviations above the mean latency for that problem type. In contrast to the results of the previous experiments, significant positive skew was evident in the distributions of the resulting mean latencies for all but the six-choice problems. Z_{skew} = 5.46 for 3-choice problems, 4.27 for 4-choice problems, and 2.44 for 5-choice problems, all p < .015. (Some also showed significant kurtosis.) The mean latencies are summarized in Table 13. To correct for the skew in the mean latencies, a log transformation was applied to them for all problem sizes. The distributions of the transformed latencies did not show significant deviation from normality, $|Z_{\text{skew}}| < 1.57$ for every problem size. I report analyses of the transformed latencies, except when comparing results with the previous experiments. For the latter, I report analyses of untransformed mean latencies.

As in the previous two experiments, children tended to take longer to map a novel label correctly than to map a familiar label correctly. This difference was significant for every problem size, all ts > 4.51, p < .001 (df varied with problem size). Just like in the first two experiments, mean latencies to map novel labels correctly were much more variable from one child to another than mean latencies to map familiar labels were.

A 4 (problem size: 3 vs. 4 vs. 5 vs. 6 choices) repeated measures ANOVA of the log-transformed mean latencies for correct responses on the novel label trials was conducted. This analysis only included 27 children because the other children lacked correct responses for at least one of the problem sizes. Because Mauchly's test was not significant, χ^2 (5) = 3.77, $p = .58$, sphericity was assumed. The main effect of problem size was significant, F (3, 78) = 8.45, $p < .001$. As expected and as in the previous two experiments, mean latency increased as the number of choices in the problem increased, F (1, 26) = 27.88, $p < .001$, $\eta_p^2 = .52$, for the linear contrast. As can be seen in Table 13, (untransformed) mean latency increased 0.51 sec on average for every increment in problem size (i.e., with each addition of one choice), which clearly differs from the mean latency of responses to familiar labels, which increased only 0.09 sec as the number of choices increased.

The overall linear trend in the latencies mirrored the linear trend in the accuracy of novel label mapping. As problem size increased, children chose the unfamiliar object less frequently and took more time before selecting it. This result replicates the linear trends in accuracy and latency observed in Experiment 2, as well as in Experiment 1 for problems involving six or fewer choices.

Table 15 summarizes correlations between various predictors and log-transformed mean latencies of correct response on novel label trials. In contrast to Experiment 2, none of these correlations between a predictor and overall mean was statistically significant. It is unlikely that this non-replication is completely due to sampling error. For example, the correlation between object nameability judgment and overall mean response latency on novel labels trials was significantly lower in Experiment 2 (r = -.64) than in Experiment 3 (r = -.20),$z_{difference}$ = 1,96, p = .050.

As I argued regarding the weaker relation between age and response latencies for familiar labels in Experiment 3, the change in procedure from a pointing response to a verbal response (i.e., "I found it.")

may be at least part of the explanation. The age-related variance in speed to point to an object may be greater than that in speed to say "I found it." Venetsanou et al. (2009)

<u>*b) Latency of incorrect responses.*</u> Analyses of latencies for incorrect responses on novel label trials included all problem-size types. Twenty, 14, 23, and 30 children responded incorrectly at least once on 3-, 4-, 5-, and 6-choice problems, respectively. Unlike in the previous experiments, mean latencies were not significantly different for correct responses and incorrect responses for any problem size. For 3-choice problems, *M* = 3.35 sec (*SD* = 2.14) as compared to *M* = 5.05 sec (*SD* = 5.28), *t* (18) = 1.94, *p* = .068; for 4-choice problems, *M* = 2.59 sec (*SD* = 1.15) as compared to *M* = 2.81 sec (*SD* = 1.31), *t* (11) = 0.51, *p* = .619; for 5-choice problems, *M* = 3.83 sec (*SD* = 2.11) as compared to *M* = 8.20 sec (*SD* = 13.90), *t* (15) = 1.36, *p* = .195; for 6-choice problems, *M* = 4.743 sec (*SD* = 2.17) as compared to *M* = 5.95 sec (*SD* = 6.86), *t* (21) = 1.13, *p* = .27. Because only 12 to 22 children had at least one correct and one incorrect response on a problem of the same size, these comparisons lacked statistical power.

The mean latency for incorrect responses actually decreased by 0.86 sec from 3- to 4-choice problems, though this difference was not statistically significant, *t* (9) = 0.53, *p* = .602. The mean latency then increased by 4.39 sec from 4- to 5-choice problems, which also was not a statistically significant difference, *t* (12) = 1.04, *p* = .304. Finally, mean latency for incorrect responses decreased by 1.54 sec from 5- to 6-choice problems, but this difference too was not statistically significant, *t* (20) = 2.24, *p* = .04. The difference in mean latency between 4- and 6-choice problems was also significant, *t* (12) = 0.61, *p* = .542.

Figure 12 depicts the relation between individuals' latencies for incorrect and correct novel label responses on each problem type. (Outliers were not corrected for these figures.). For three-choice problems, the latencies of the two types of responses were strongly correlated, *r* (17) = .70, *p* = .001. However, this was the only problem size for which the correlations between correct and incorrect latencies was statistically significant by a two-tailed test. For four-choice problems, the latencies of the two types of responses were not significantly correlated, *r* (10) = .32, *p* = .309. For five- and six-choice problems, the correlations were significant by one-tailed tests, r (14) = .46, one-tailed p = .072, and r (20) = .40, one-tailed p = .069, respectively. Overall across problem sizes, mean correct latency was strongly correlated with mean incorrect latency, r (31) = .77, p < .001. Thus, as in the previous two experiments, there was evidence that the children who tended to take longer to make correct selections on novel label trials also tended to take longer to make incorrect selections.

Analyses of eye-tracking measures. Using Tobii Pro Studio software, I analyzed three dependent variables with respect to the tracking of children's gaze patterns. The first variable was the total number of foils checked on each trial. For each trial, an area of interest (AOI) was drawn around each object in the array. These AOIs were categorized as either a target or a foil. Then, for each individual trial, the Tobii software calculated the number of individual fixations within a given AOI and differentiated according to these categories. This particular variable is a count of the number of foils on which a child fixated at least once from the appearance of the array until the child indicated they had found the referent.

The second dependent variable I analyzed was the number of revisits to the target, that is, the number of times a child returned to fixate on the target after already having fixated on it once for a given trial. For example, a child who heard a novel label, fixated on the target object, then fixated on a distractor or two, and then returned to fixate on the target object before verbally declaring he or she had found the target object was given a value of 1 for that item, since there was one revisit to the target. On the contrary, a child who heard a novel label, fixated on the target, and then made a selection without revisiting, was given a value of 0, since there were no revisits to the target. This variable was coded manually by several research assistants who watched playback of each child's gaze patterns for each trial.

The final variable I analyzed was the proportion of time spent looking at the target relative to time spent looking at the familiar object foils. Tobii software generates total time spent fixating within a given AOI. The measure was calculated by dividing the fixation time spent in the target's AOI by the sum of fixation time spent in the target's AOI and total fixation time spent in the AOIs of competitors. This measure was set up so that a value closer to 1 indicates more relative time fixating on the target while a value closer to 0 indicates more relative time fixating on the foils. For example, a child who spent 500 ms total fixating on the target object and 100 ms total fixating on the only foil he ever checked was given a value of .83 (500 / (500+100)), whereas a child who spent 250 ms total fixating on the target object and 900 ms total fixating on some of the foils received a score of .22 (250 / (250 + 900)).

Relations between eye-tracking measures and novel-name mapping accuracy. The associations between each of the three eye-tracking measures and disambiguation accuracy are reported below. The distributions and means of these measures were examined for outliers and normality. Outliers were defined as values that were more than three standard deviations above or below the mean.

a) Number of foils checked. Two outliers were identified, one occurring on a 4-choice problem and one on a 6-choice problem. Each of these outliers was replaced by a value 2.33 standard deviations above the mean. The resulting distributions did not deviate from normality.

A 4 (problem size) repeated measures ANOVA of these data was conducted. Mauchly's test indicated that the assumption of sphericity had been violated, χ^2 (5) = 18.53, p = .002, so degrees of freedom were corrected using the Greenhouse-Geisser estimate of sphericity (ε = .72). There was a significant effect of problem size, F (1.96, 43.10) = 30.87, p < .001. Only the linear trend component of this effect was significant, F (1, 22) = 130.65, p < .001, η_p^2 = .86. As number of choices increases, so did the mean number of foils checked.

The correlation between mean number of foils checked and the mean of the accuracy for each problem size was significant, r (35) = .53, p = .001. Children who tended to check more foils on novel label trials tended to be more accurate. This relation was evident for 5-choice (r = .39, p = .016) and 6-choice problems (r = .38, p = .019) and trended in the predicted direction for 4-choice problems (r = .31, p = .062). There was no association between foils checked and accuracy on 3-choice problems, r = -.04, p = .826. The rate at which number of foils checked increased as problem size increased was not correlated with the rate at which accuracy decreased as problem size increased, r = .26, p = .125.

Table 16 shows the correlations between the various predictors and foils checked on novel label trials. Age and receptive vocabulary were both positively correlated with the number of foils checked. Older children and children with greater verbal intelligence were likely to check more foils. There was no significant correlation observed between foils checked and awareness of lexical knowledge, however. Thus, there was no evidence that the significant relation between awareness of lexical knowledge and overall disambiguation accuracy was even partially due to number of foils checked. That is, high-aware children tended to map novel labels more accurately than low-aware children even though they tended to check as many foils as the low-aware children. The correlation between awareness of lexical knowledge and overall mean accuracy was significant when overall mean number of foils checked was partialled out, r (34) = .36, p = .032.

Interestingly, children consistently disambiguated correctly without first checking all of the familiar object foils. They fixated on significantly fewer than the total number of foils on correct trials for every problem size. On 3-choice items (i.e., 2 foils), the mean number of foils checked was 1.49 (*SD* = .60), t (37) = 5.28, $p < .001$. On 4-choice items, the mean number of foils checked was 1.88 (*SD* = .76), t (35) = 8.80, $p < .001$. On 5-choice items, children fixated on an average of 2.52 foils (*SD* = 1.10), t (30) = 7.50, $p < .001$, and on 6-choice items, children fixated on an average of 3.73 foils (*SD* = 1.21), t (29) = 5.72, $p < .001$. The finding that children regularly choose the novel object as the referent of a novel label without first checking (and eliminating) all of the familiar object foils was surprising when considering that most of the leading accounts of the disambiguation effect involve first rejecting the familiar object(s) (Markman, 1989; Diesendruck & Markson, 2001) or engaging in disjunctive syllogism (i.e., process-of-elimination) (Halberda, 2006).

One explanation for this finding is that a child represents the novel label as a word that they do not know and then chooses the novel object because it is also one they do not know (Slocum & Merriman, 2018; Henning & Merriman, 2019). Although awareness of lexical knowledge was not correlated with foils checked overall, there is some evidence that checking more foils before selecting an object is associated with greater accuracy for children with lesser awareness of their lexical knowledge. In low-aware children, the number of foils checked was significantly correlated with disambiguation accuracy, r (15) = .69, p = .002. In contrast, for high-aware children, this correlation was weaker, r (18) = .29, p = .22. This difference in correlation coefficients between high- and low-aware children was approaching statistical significance, $z_{difference}$ = 1.69, p = .09.

These results suggest that for children who lack awareness of their lexical knowledge, it may be important to check more of the foils before selecting. The more foils these children checked, the more likely they were to be accurate. In contrast, checking more foils does not necessarily lead to greater accuracy for children who possess an awareness of their own knowledge. These children were no more likely to be accurate whether they checked few or many foils first. I suggest that these children notice that the label is a word they do not know, and they search for the object that matches metacognitively. That is, they search for the object that they do not know.

b) Number of revisits. No outliers were identified, and the distributions of mean revisits for each problem size did not deviate from normality.

A 4 (problem size) repeated measures ANOVA of these data was conducted. Mauchly's test indicated that the assumption of sphericity had not been violated, χ^2 (5) = 0.56, p = .990, so sphericity was assumed. There was a significant effect of problem size, F (3, 69) = 3.25, p = .027. Only the quadratic trend component of this effect was significant, F (1, 23) = 10.42, p = .004, η_p^2 = .31. Children tended to make few revisits on 3- and 6-choice trials, and they tended to revisit the target significantly more frequently on 4- and 5-choice trials.

The correlation between mean number of revisits and the mean of the accuracy for each problem size was significant by a one-tailed test, r (35) = .32, one-tailed p = .055. Children who tended to revisit the target object before choosing on novel label trials tended to be more accurate. This relation was only evident for 6-choice problems (r = .40, p = .013) and approached statistical significance in 4-choice (r =

.30, p = .077) and 5-choice problems (r = .29, p = .082). There was no association between revisits and accuracy on 3-choice problems, r = -.06, p = .724.

Table 17 shows the correlations between the various predictors and number of revisits on novel label trials. Once again, age and receptive vocabulary were both positively correlated with the number of revisits, indicating that older children and children with greater verbal intelligence were likely to revisit the target. Once again there was no significant correlation observed between revisits and awareness of lexical knowledge. Thus, like with foils checked, there was no evidence that the significant relation between awareness of lexical knowledge and overall disambiguation accuracy was even partially due to number of revisits. In other words, high-aware children tended to map novel labels more accurately than low-aware children even though they tended to revisit the target as frequently as the low-aware children. The correlation between awareness of lexical knowledge and overall mean accuracy was significant when overall mean number of revisits was partialled out, r (34) = .39, p = .019.

Like with foils checked, in children who did not demonstrate an awareness of their lexical knowledge, the number of revisits was positively correlated with accuracy, r (16) = .54, p = .021. Also like with foils checked, there was no significant correlation between number of revisits and accuracy for high-aware children, r (18) = .16, p = .512. However, with respect to revisits, the difference between these two correlation coefficients was not statistically significant, $z_{difference}$ = 1.25, p = .211. Interestingly, number of revisits was positively correlated with number of foils checked, r (37) = .66, p < .001, which is the case in both low-aware (r = .56, p = .01) and high-aware children (r = .79, p < .001). The correlation remains even after controlling for age and awareness of lexical knowledge, partial r (34) = .60, p < .001. This suggests that revisits and foils checked are linked in some way, perhaps that some children are more careful than others. The results from the current experiment indicate that if that is the case, children without awareness of their lexical knowledge may need to be careful in order to be accurate, whereas carefulness is may not be as important for children who do possess this kind of awareness.

c) Proportion of time spent looking at the target. No outliers were identified, and the distributions of the proportion of looking time for each problem size did not deviate from normality.

To explore the effect of problem size on proportion of looking time on the target for the critical novel label trials, an additional 4 (problem size: 3 vs. 4 vs. 5 vs. 6 choices) repeated measures ANOVA was conducted. Once again, Mauchly's test was not significant, χ^2 (5) = 2.47, $p = .782$, so sphericity was assumed. The main effect of problem size was significant, F (3, 105) = 35.14, $p < .001$. As with familiar label trials, the mean proportion of time spent looking at the target on novel label trials decreased as the number of choices in the problem increased, F (1, 35) = 122.41, $p < .001$, $\eta_p^2 = .78$, for the linear contrast as well as the quadratic contrast, F (1, 35) = 9.90, $p = .003$, $\eta_p^2 = .22$. Mean proportion of relative looking time on the target remained about the same from 3- to 4-choice problems, and then decreased linearly as the number of choices increased.

The correlation between proportion of looking time on the target relative to the foils and mean of the accuracy for each problem size was significant by a one-tailed test, r (35) = .31, one-tailed $p = .064$. This relation was evident for all problem sizes, rs between .40 and .58, $p < .02$, except for 4-choice problems, $r = .12$, $p = .471$. The correlation observed was even stronger when examining the looking just before a selection is made. In the 500 milliseconds before a child chooses an object, relative proportion of looking time spent on the target was associated with greater accuracy, r (35) = .51, $p < .001$. Children who tended to spend more of their time looking at the target object, especially immediately before choosing, tended to choose accurately.

Table 18 summarizes correlations between various predictors and mean proportion of looking time spent on the target for novel label trials. Age was the only predictor that was significantly correlated with proportion of looking time on the target. Older children tended to spend more of their time fixated on the target object. Once again, there was no significant correlation observed between this measure and awareness of lexical knowledge. Therefore, like with the other two eye-tracking measures, there was no evidence that the significant relation between awareness of lexical knowledge and disambiguation accuracy was even partially due to the percentage of the time they spent attending to the target. The correlation between awareness of lexical knowledge and overall mean accuracy remained significant after partialling out the proportion of looking time on the target, r (34) = .40, p = .016.

Relations between eye-tracking measures and latency for familiar label trials. The distributions and means for the eye-tracking measures and familiar label mapping latencies have already been examined for outliers and normality. The same Winsorized and log-transformed values which have been used for previous analyses are also used here. The results for each of the measures are reported separately below.

a) Number of foils checked. Overall, children rarely fixated on many objects other than the target. The mean number of foils checked for all problem sizes was 1.11 (*SD* = .07). In other words, children never checked (i.e., ignored) an average of 2.39 objects (or 68.2%) on familiar label trials. Broken down by problem size, children never checked an average of 62.5% of foils for 3-choice items, 68.3% for 4-choice items, 65% for 5-choice items, and 73.4% for 6-choice items. These results suggest that hearing a known word on a familiar label trial serves as a cue for children to search for the visual representation of that label's referent and find it rapidly.

A 4 (problem size) repeated measures ANOVA of these data was conducted. Since Mauchly's test was not significant, $\chi 2$ (5) = 7.72, p = .172, sphericity was assumed. There was a significant effect of problem size, F (3, 108) = 11.86, $p < .001$. Only the linear trend component of this effect was significant, F (1, 36) = 25.65, $p < .001$, η_p^2 = .403. As number of choices increases, so did the mean number of foils checked. Interestingly, children tended to check only about 35% of the foils on each of the four sizes of problems. If one assumes that once children fixate on the correct referent of a familiar label they do not continue to check other objects, then they should check 50% of the foils on average on every size of problem. This percentage may be somewhat lower if they are able to identify some familiar objects when fixating near, but not in the object's AOI.

The correlation between mean number of foils checked and the mean of the log-transformed reaction times for each problem size was significant, r (35) = .50, $p < .001$. Children who took longer to respond on familiar labels trials tended to check more foils. This relation was evident for each size of problem, *r*s ranged from .56 to .57, $p < .001$, except for 3-choice problems, where $r = .15$, $p = .38$. The rate at which number of foils checked increased as problem size increased was also correlated with the rate at which log-transformed reaction time increased as problem size increased, $r = .58$, $p < .001$.

Despite this correlation, the overall mean number of foils checked on familiar label problems was not significantly correlated with any of the predictors (see Table 19). Thus, there was no evidence that the significant relation between awareness of lexical knowledge and overall latency to map familiar labels was even partially due to number of foils checked. That is, high-aware children tended to map familiar labels faster than low-aware children even though they tended to check as many foils as the low-aware children. The correlation between awareness of lexical knowledge and overall mean latency to map familiar labels was significant (by a one-tailed test) when overall mean number of foils checked was partialled out, r (34) = -.31, one-tailed $p = .065$.

b) Number of revisits. Overall, revisits were very rare. The mean number of revisits for all problem sizes was 0.17 (*SD* = .20). Broken down by problem size, a single revisit only occurred on 14% of all possible 3-choice trials, 15% of 4-choice trials, 19% of 5-choice trials, and 22% of 6-choice trials. This suggests that for familiar label trials, not surprisingly, once a child fixates on the target object once, they almost never need to return before correctly choosing the target object as the referent of the label.

A 4 (problem size) repeated measures ANOVA of these data was conducted. Since Mauchly's test was not significant, $\chi2$ (5) = 9.27, $p = .10$, sphericity was assumed. There was no significant effect of problem size, F (3, 108) = 1.79, $p = .153$. However, the linear trend component of this effect was significant, F (1, 36) = 4.62, $p = .038$, $\eta_p^2 = .114$. On average, there were .13 (*SD* = .20) revisits on 3-choice trials, .14 (*SD* = .30) on 4-choice trials, .18 (*SD* = .24) on 5-choice trials, and .24 (*SD* = .34) on 6-

choice trials. Revisits did increase in a linear fashion as the number of foils increased, but there were very few in general.

The correlation between mean number of revisits and the mean of the log-transformed reaction times for each problem size was significant, *r* (35) = .367, *p* = .025. Children who took longer to respond on familiar labels trials tended to revisit the target more often. This relation was evident for each size of problem, *r*s ranged from .33 to .46, $p < .05$, except for 3-choice problems, where *r* = .21, *p* = .212. The rate at which number of revisits increased as problem size increased was also correlated with the rate at which log-transformed reaction time increased as problem size increased, *r* = .45, *p* = .001.

Despite this correlation, the overall mean number of revisits on familiar label problems was not significantly correlated with any of the predictors (see Table 20). Thus, there was no evidence that the significant relation between awareness of lexical knowledge and overall latency to map familiar labels was even partially due to number of revisits. Once again, high-aware children tended to map familiar labels faster than low-aware children even though they tended to revisit the target as frequently as the low-aware children. The correlation between awareness of lexical knowledge and overall mean latency to map familiar labels was significant overall mean number of revisits was partialled out, *r* (34) = -.38, *p* = .024.

c) Proportion of time spent looking at the target. A 4 (problem size) repeated measures ANOVA of mean proportion of looking time on the target, relative to the foils was conducted. Because Mauchly's test was not significant, χ^2 (5) = 3.45, *p* = .631, sphericity was assumed. The main effect of problem size was significant, *F* (3, 108) = 7.89, $p < .001$. Both the linear, *F* (1, 36) = 13.54, *p* = .001, η_p^2 = .27, and quadratic, *F* (1, 36) = 6.94, *p* = .012, η_p^2 = .16, components of this effect were significant. The relative proportion of time spent fixating on the target decreased from 3- to 5-choices, where it remained about the same for 6-choice problems.

The correlation between mean proportion of relative looking time to the target and the mean of the log-transformed reaction times for each problem size was significant, *r* (35) = -.40, *p* = .015. Children who responded more quickly tended to spend more of their time looking at the target object relative to the foils. This relation was evident for each size of the problem, *r*s ranged from -.32 to -.41, *p* = .05, except for 3-choice problems, where *r* = -.04, *p* = .81. The rate at which proportion of looking time on the target increased as problem size increased was also negatively correlated with the rate at which log-transformed reaction time increased as problem size increased, *r* = -.37, *p* = .024.

Despite this correlation, the overall mean proportion of looking time on the target for familiar label problems was not significantly correlated with any of the predictors (see Table 21). Thus, there was no evidence that the significant relation between awareness of lexical knowledge and overall latency to map familiar names was even partially due to looking time to the target. Once again high-aware children tended to map familiar names faster than low-aware children even though they tended to spend as much relative time looking at the target as the low-aware children. The correlation between awareness of lexical knowledge and overall mean latency to map familiar labels was significant when overall mean proportion of looking time was partialled out, *r* (34) = -.37, *p* = .025.

Relations between eye-tracking measures and latency for novel label trials. The distributions and means for the eye-tracking measures and novel label mapping latencies have already been examined for outliers and normality. The same Winsorized and log-transformed values which have been used for previous analyses are also used here. The results for each of the measures are reported separately below.

a) Number of foils checked. The correlation between mean number of foils checked and the mean of the log-transformed reaction times for each problem size was significant, *r* (23) = .63, *p* = .001. Not surprisingly, children who tended to check more foils also tended to take longer to respond on correct novel labels trials. This relation was evident for each size of problem, *r*s ranged from .38 to .77, *p* < .03. The rate at which number of foils checked increased as problem size increased was also correlated with the rate at which log-transformed reaction time increased as problem size increased, *r* = .68, *p* < .001.

As can be seen in Table 16, the only predictors that were associated with foils checked for novel label trials were age and receptive vocabulary. There was no relationship observed between foils checked and awareness of lexical knowledge. Older children and children with greater verbal intelligence were likely to check more foils before choosing the novel object as the referent of the novel label. High-aware children, however, were no more likely than low-aware children to check fewer foils.

b) Number of revisits. Similar to foils checked, the correlation between mean number of revisits and the mean of the log-transformed reaction times for each problem size was significant, r (23) = .65, p = .001. Predictably, children who tended to revisit the target before selecting also tended to take longer to respond correctly on novel label trials. This relation was evident for each size of the problem, rs ranged from .50 to .69, $p < .01$. The rate at which number of revisits increased was also correlated with the rate at which log-transformed reaction times increased as problem size increased, $r = .46$, $p = .023$.

As can be seen in Table 17, the only predictors that were associated with revisits for correct novel label trials were receptive vocabulary [and age at the one-tailed level]. There was no evidence of a relationship between awareness of lexical knowledge and number of revisits. Children with greater verbal intelligence tended to revisit the target before choosing more frequently. The finding that these children tended to check more foils and revisit more often makes sense given the task used to assess verbal intelligence. I used the Peabody Picture Vocabulary Test (PPVT) (Dunn & Dunn, 2007) as the measure of receptive vocabulary, or verbal intelligence. In this task, children hear a word and must choose from a group of four pictures, which one is the referent of this word. As the PPVT advances and becomes more difficult, the pictures become more similar to one another, and carefulness often promotes success. Children are more likely to succeed in both tasks if they are sure to check all of the possible options and double-check when the choice is not clear.

c) Proportion of time spent looking at the target. There was no significant correlation observed between mean proportion of time spent looking at the target relative to foils and the mean of the log-transformed reaction times for each problem size, r (23) = -.14, p = .51. In addition, there was no significant correlation observed between these two variables for any problem size, *r*s ranged from -.24 to +.19, $p > .2$. Also, the rate at which relative proportion of looking time on the target increased as problem size increased was not significantly correlated with the rate at which log-transformed reaction time increased as problem size increased, $r = -.08$, $p = .69$. Unlike familiar label trials, children who tended to spend more of their time looking at the target did not tend to choose the target object more quickly.

As can be seen in Table 18, the only predictor that was significantly associated with proportion of looking time spent on the target was age in months. Older children tended to spend more of their time fixating on the target compared to the foils than younger children. Once again, awareness of lexical knowledge was not associated with proportion of time spent on the target.

General Discussion

The three experiments in this dissertation tested the hypothesis that children with awareness of their own lexical knowledge solve disambiguation problems more accurately than children who lack this awareness, and that this advantage increases as the number of distractors in the problem increases. In every experiment, disambiguation accuracy decreased linearly as the number of distractors increased. Additionally, in every experiment, children who demonstrated an awareness of their lexical knowledge disambiguated more accurately overall than children who failed to demonstrate awareness of their own knowledge. In Experiments 1 and 2, awareness of lexical knowledge was also associated with a less severe decline in accuracy as the number of distractors increased up to six total objects. When the disambiguation task became more demanding, awareness of lexical knowledge appeared to play a key role in allowing young children to continue to map a novel label to the one novel object among foils.

The experiments in this dissertation also examined the effect of disambiguation problem size on latency of correct responses. Mirroring the trend in accuracy, latency increased linearly as the number of distractors increased up to six total objects. Also similar to accuracy, awareness of lexical knowledge predicted a less severe latency increase as the number of distractors increased in the two experiments that included up to six objects. While awareness of lexical knowledge only predicted faster overall response latencies in Experiment 2, the addition of distractors had a larger effect on children who lacked awareness of their lexical knowledge than children who possessed this awareness in two of three experiments.

Eye-tracking was used to explore the nature of the relationship of awareness of lexical knowledge with accuracy and latency in Experiment 3. There was no evidence in this experiment that either the total number of foils checked or the number of revisits could account for these relationships, as awareness of lexical knowledge was not associated with either of these eye-tracking measures. These measures did, however, shed some light on how important checking more foils and revisiting the target object are for children who lack awareness of their lexical knowledge. For these children, disambiguation accuracy was strongly correlated with checking more foils and revisiting the target more frequently. On the other hand, disambiguation accuracy was not correlated with either of these measures for children who showed an awareness of their knowledge. Finally, Experiment 3 provided evidence that relative time spent fixating on the target relative to the distractors is strongly associated with accuracy. Tracking children's eye movements showed that the more time a child spends looks at the target object, the more likely he or she is to select it as the referent of a novel label.

The finding that awareness of lexical knowledge predicts an increasing accuracy advantage as the number of distractors increases suggests that metacognitive development that takes place around the fourth birthday plays a role in children's ability to disambiguate. Previous work has suggested that children who demonstrate this kind of metacognitive awareness can represent labels and objects as "ones I know" and "ones I don't know" (Slocum & Merriman, 2018; Henning & Merriman, 2019). I propose that these metacognitions become especially important as the number of distractors in a disambiguation problem increases. In a simple disambiguation problem, in which a child must decide between one familiar object and one novel object, metacognition may not be necessary. A child can pass this task consistently by retrieving the label for the familiar object, noting that it mismatches the novel label presented, and picking the novel object because of the mismatch. However, with several familiar object foils present (e.g., five), the likelihood of successfully carrying out these numerous processes decreases. The child with more advanced metacognitive awareness is aided by his or her ability to recognize that the novel label is unfamiliar and so is only one of the objects.

Support for this proposal comes from the finding that while there was a significant difference in accuracy, there was no difference in the number of foils on which high- and low-aware children fixated. Thus, among children who tended to check the same number of foils on novel label problems, those with advanced metacognitive awareness mapped the label to the novel object more reliably. One possible explanation is that when a child noticed that the novel object and novel label were both ones they did not know, they tended to select the novel object even if they had not yet checked several of the foils. Conversely, when a child did not notice this metacognitive match, they tended to select the novel object only if they had checked and rejected most or all of the foils. This conclusion fits with previous work in which only children with advanced awareness of their lexical knowledge succeeded in a disambiguation task in which label retrieval was not possible (Slocum & Merriman, 2018).

However, there is a second possibility. Children with advanced metacognition may just have been more likely than other children to reject the foils that they did check, which could have promoted their mapping the novel label to the novel object. Their metacognitive realization that these were objects that they already knew may have promoted their rejection of them. Also, they may just have retrieved the familiar label for a foil and noted its mismatch with the novel label more reliably. Children who make more accurate judgments of whether they know what a word means, or whether they know the name for an object, also tend to execute non-metacognitive processes such as object recognition, label retrieval, and word meaning retrieval more rapidly and reliably (Lipowski & Merriman, 2011; Merriman & Lipko, 2008).

The two explanations just discussed can also be applied to the finding that as the number of foils increased, children with greater awareness of lexical knowledge showed a more gradual increase in latency to correctly map the novel label to the novel object. On the one hand, their ability to reflect on the feeling of novelty that the label evoked may have promoted mapping it to the novel object more quickly, especially on trials in which many distractors were present. An increase in foils would not lead to as great an increase in latency for these children, compared to children who relied primarily on retrieving labels for the familiar objects and rejecting those objects as the intended referent. The other explanation, however, is that the metacognitively-advanced children just rejected the foils that they did check more rapidly because they realized that these objects were ones they already knew and/or more rapidly executed non-metacognitive processes that supported rejecting the objects (e.g., retrieving its familiar name).

If the first explanation were correct, children with advanced metacognitive awareness ought to check fewer familiar object foils than the less advanced children, especially for the larger problem sizes.

The difference between how many foils these two groups of children check (on trials in which they ultimately select the novel label) ought to increase as number of foils increases. However, there was no evidence of a relationship between awareness of lexical knowledge and total number of familiar object foils checked. In other words, there was no evidence that high-aware children fixated on fewer foils before choosing the target than low-aware children did. Thus, there is more support for the second explanation.

More recent studies on the disambiguation effect have focused on the distinction between fast-mapping of a novel label onto a novel object and later retention of that mapping (Horst et al., 2010; Horst & Samuelson, 2008; Bloom, 2000). Experiment 2 included a retention task like the one used in these studies and, importantly, provided evidence that disambiguation retention was positively associated with awareness of lexical knowledge. Children who were able to reflect on their own lexical knowledge were more likely to retain an initial name-object mapping after a delay. Horst et al. (2010) proposed that in order to successfully map and retain a novel label onto a novel object, a child must first attend to the distractors to reject them as potential referents, and then switch their attention to the target in order to encode something about the novel object, so that it may be recalled later. Perhaps metacognitive awareness allows the initial mapping of a novel label onto a novel object to be more salient. For example, a child without this kind of reflective awareness may reject several competitors before choosing the novel by elimination. However, a more metacognitively advanced child may recognize the label (e.g., *blicket*) as "one I don't know" and engage in a more active search for a novel object. It is possible that the latter search pattern leads to greater encoding of the properties of the novel object and a greater likelihood of recalling the mapping later.

Another finding from Horst et al. (2010), which was replicated here, was that children retained more mappings than would be expected by chance on initial three-choice trials, but not when the number of choices exceeded three. They suggested that as the number of competitors increases, children spend a higher proportion of their time attending to the wrong objects. In other words, with more distractors, children spend less relative time looking at the target as compared to the competitors during the initial mapping phase. However, Horst et al. (2010) did not measure eye-tracking; this hypothesis was based on latency and accuracy data. The current study provided evidence for their suggestion, showing a main effect of number of competitors, such that as the problem gets larger, the amount of relative looking to the target decreases linearly. The current study also provided evidence that attention on the target predicts greater accuracy of the initial name-object mapping, but future work should examine the relationship between relative looking time and accuracy of retention, in order to determine whether more relative looking to the target is directly or indirectly related to retention of a mapping, as well as whether awareness of lexical knowledge moderates the relationship.

Taken together, the experiments in the current investigation provided evidence that children with greater awareness of their lexical knowledge exhibit a shallower rate of decline in accuracy as well as speed of disambiguation as the problem size increases, and they are more likely to retain the initial name-object mapping after a delay. These general results fit with much of the previous work on the role of metacognition in disambiguation of novel name reference. For example, previous research has shown the accuracy of lexical knowledge judgments to be positively associated with the strength of the disambiguation effect in a standard paradigm (Merriman & Bowman, 1989; Merriman & Schuster, 1991) as well as a cross-modal paradigm (Wall et al., 2015). In addition, Marazita and Merriman (2004) found that 4-year-olds who made accurate judgments of their lexical knowledge were more likely to offer a disambiguation justification, which acknowledged the rejection of the familiar object. Finally, two recent studies demonstrated that only those children who made accurate judgments of their lexical knowledge were capable of disambiguating metacognitively, that is, without visual access to the familiar object (Slocum & Merriman, 2018) and without auditory access to the label (Henning & Merriman, 2019).

The results from the current investigation also fit with an abundant literature showing advances in various aspects of children's understanding of knowledge and mental states around the transition from three to four years of age. Children use words like *know* and *think* in conversation much more frequently to make reference to belief states (Bartsch & Wellman, 1995; Moore, Furrow, Chiasson, & Patriquin, 1994; Hughes & Dunn, 1998) about the same time that they come to understand that “knowing (that)” denotes more certainty than “thinking (that)” (Johnson & Maratsos, 1977; Moore, Bryant, & Furrow, 1989; Cherney, 2003). During this period, children develop the understanding that experiencing a stimulus through a different sensory modality (e.g., touching it versus looking at it) causes someone to acquire different kinds of knowledge about it (e.g., its texture versus its color) (O’Neill, Astington, & Flavell, 1992).

In the third experiment, analysis of children’s gaze patterns indicated that they often disambiguated accurately without first checking all of the distractors. This finding was relatively surprising, in particular with respect to the existing accounts of the disambiguation effect in young children. Both the Mutual Exclusivity (Markman & Wachtel, 1989; Merriman & Marazita, 1995) and Pragmatic Contrast accounts (Clark, 1990; Diesendruck & Markson, 2001; Gathercole, 1989) suggest that children ultimately disambiguate based on a mismatch between a representation of the novel label and a representation of each of the objects from which to choose. While these accounts could easily acknowledge that these decisions sometimes involve comparing metacognitive representations as well, the eye-tracking evidence observed here suggests that children sometimes disambiguate without considering all possible objects.

More recent studies have also shown through eye-tracking and measuring response latency that preschoolers use disjunctive syllogism (i.e., process-of-elimination) to learn new words (Halberda, 2006; Horst et al., 2010). However, Halberda’s eye-tracking study only involved trials with one familiar object foil. Perhaps children utilize disjunctive syllogism when only two total objects are present, but use a different strategy when there are at least three objects from which to choose. It is possible that the results observed here fit better with something like the Novel-Name, Nameless Category (NC3) account, which suggests that children solve disambiguation problems by mapping novelty to novelty (Golinkoff, Hirsh-Pasek, Bailey, & Wenger, 1992). However, if children were mapping novelty to novelty in the current study, we ought not have observed a linear increase in response latency or foils checked as the number of distractors increased. Once again, it is possible that the strategy a child uses for a trial involving just two objects changes with the addition of familiar objects. Nevertheless, it is important that the existing accounts of the disambiguation effect are updated to consider the presence of distractors, a situation that resembles a more realistic scenario for a young child encountering early exposure to language.

One limitation of the current study concerns the relatively small sample size utilized in each experiment. While the sample sizes of the experiments here were similar to other developmental studies involving preschoolers, power analyses indicate that our sample size allows for the detection of a large effect, but not a moderate one. In the current investigation, there were several examples of effects that seemed robust by the eye test, but did not meet the threshold of statistical significance. Aside from ensuring that each successive study includes a substantially larger sample than the ones collected here, another potential solution is to pool the results of several replication studies, rather than considering studies individually, using a continuously cumulating meta-analytic (CCMA) approach (Braver, Thoemmes, & Rosenthal, 2014; Rosenthal, 1990).

Another limitation in the current investigation is that it is likely the procedural change in the third experiment was more consequential than initially anticipated. There was some evidence that eliminating the opportunity for children to hear the label and repeat it back to the experimenter before being asked to find its referent had a negative effect on disambiguation accuracy, especially for children with awareness of their own knowledge. I proposed that having the ability to hear the novel label once and repeat it gave the more advanced children a metacognitive jump start. Similar to how the presentation of a familiar label prompts a child to call up visual representations of that label's meaning, the earlier presentation of the novel label may trigger more advanced children to realize that no meaning is coming to mind and thus, the label is one they do not know. This realization may similarly encourage them to search for the object they do not know, or the one that is unfamiliar. A future study should test this hypothesis directly by including one condition in which the procedure is identical to that of the second experiment, and one condition whose procedure is identical to that of the third experiment.

Another future direction is to explore the patterns of children's eye gaze for trials on which they responded inaccurately. Specifically, a future study should examine the amount of looking at the target relative to the competitors during the window immediately preceding the verbal selection. There is a substantial literature of studies that suggest that children implicitly understand false belief in an unexpected change-of-location task (Onishi & Baillargeon, 2005; Surian, Caldi, & Sperber, 2007) and an unexpected contents task (Buttelmann, Over, Carpenter, & Tomasello, 2014), as well as understanding others' mental states (Scott, Richman, & Baillargeon, 2015) before they will indicate an explicit understanding by responding verbally (for a review, see Low & Perner, 2012). Perhaps children endure a similar transitional phase in which they understand that the novel object is the referent of the novel label, but they fail to make this choice explicitly. In this case, they may look significantly longer at the target relative to the familiar object foils, but ultimately fail to choose correctly.

This research could have significant implications for promoting children's early vocabulary development. Recently, several interventions have focused on the importance of exposure to language during toddlerhood and early childhood. These interventions stem from the so-called "30 million-word gap" coined by Hart and Risley (1995), which estimates that by age four, children in families with lower socioeconomic status hear about thirty-million fewer words than children in families with high socioeconomic status. Research has linked this lack of exposure to language with later vocabulary development, school readiness, and literacy development (for a review, see Pace, Luo, Hirsh-Pasek, & Golinkoff, 2017). In addition to the quantity of early parent-child interactions, the current interventions emphasize the quality of these interactions as well, highlighting the importance of parental sensitivity, responsiveness, and joint attention while reading books with an emphasis on asking children open-ended questions (Pace et al., 2017; Valdez-Menchaca & Whitehurst, 1992). Even more relevant to the current study, recent work has shown that children with parents who use mental state language (e.g., think, know, understand) are more likely to demonstrate metacognitive awareness, by passing a nonverbal false-belief task, at two-and-a-half years of age (Roby & Scott, 2018). The relationship observed between another type of metacognitive awareness, namely awareness of lexical knowledge, and efficiency of word-learning in the current study suggests that early interventions may continue to stress the quality of parent-child verbal interaction, but specifically encourage the use of mental state language and references to awareness of lexical knowledge.

References

Aguiar, N. R., Stoess, C. J., & Taylor, M. (2012). The development of children’s ability to fill gaps in their knowledge by consulting experts. *Child Development, 83*, 1368-1381. doi:10.1111/j.1467-8624.2012.01782.x1

Baldwin, Dare A. (1993). Infants' ability to consult the speaker for clues to word reference. *Journal of Child Language, 20*, 395-418. https://doi.org/10.1017/S0305000900008345

Bartsch, K., & Wellman, H. M. (1995). *Children talk about the mind.* Oxford university press.

Bedford, R., Gliga, T., Frame, K., Hudry, K., Chandler, S., Johnson, M. H., Charman, T., & The Basis Team. (2013). Failure to learn from feedback underlies word learning difficulties in toddlers at risk for autism. *Journal of Child Language*, *40*(1), 29-46. https://doi.org/10.1017/S0305000912000086

Beverly, B., & Estis, J. (2003). Fast mapping deficits during disambiguation in children with specific language impairment. *Journal of Speech Language Pathology and Audiology, 27*, 163–171.

Bloom, P. (2000). How children learn the meanings of words: Learning, development and conceptual change. MIT Press.

Braver, S. L., Thoemmes, F. J., & Rosenthal, R. (2014). Continuously cumulating meta-analysis and replicability. *Perspectives on Psychological Science*, *9*(3), 333-342. doi: https://doi.org/10.1177%2F1745691614529796

Buttelmann, D., Over, H., Carpenter, M., & Tomasello, M. (2014). Eighteen-month-olds understand false beliefs in an unexpected-contents task. *Journal of experimental child psychology*, *119*, 120-126. doi: https://doi.org/10.1016/j.jecp.2013.10.002

Byers-Heinlein, K., & Werker, J. F. (2009). Monolingual, bilingual, trilingual: Infants' language experience influences the development of a word-learning heuristic. *Developmental Science*, *12*(5), 815–823. https://doi.org/10.1111/j.1467-7687.2009.00902.x

Byers-Heinlein, K., & Werker, J. F. (2013). Lexicon structure and the disambiguation of novel words: Evidence from bilingual infants. *Cognition*, *128*(3), 407–416. https://doi.org/10.1016/j.cognition.2013.05.010

Chaney, C. (1992). Language development, metalinguistic skills, and print awareness in 3-year-old children. *Applied Psycholinguistics, 13*, 485–514. doi: 10.1017/S0142716400005774

Cherney, I. D. (2003). Young children's spontaneous utterances of mental terms and the accuracy of their memory behaviors: A different methodological approach. *Infant and Child Development: An International Journal of Research and Practice*, *12*(1), 89-105. doi: https://doi.org/10.1002/icd.267

Choi, I.-R., & Hwang, M. (2014). Korean late-talkers' use of the mutual exclusivity assumption on first versus second label learning. *Communication Sciences & Disorders*, *19*(3), 285–293. https://doi.org/10.12963/csd.14169

Clark, E. V. (1990). On the pragmatics of contrast. *Journal of Child Language, 17*, 417–431. doi:10.1017/S0305000900013842

Davidson, D., Jergovic, D., Imami, Z., & Theodos, V. (1997). Monolingual and bilingual children's use of the mutual exclusivity constraint. *Journal of Child Language*, *24*, 3–24. https://doi.org/10.1017/s0305000996002917

de Marchena, A., Eigsti, I., Worek, A., Ono, K., & Snedeker, J. (2011). Mutual exclusivity in autism spectrum disorders: Testing the pragmatic hypothesis. *Cognition*, *119*(1), 96–113. https://doi.org/10.1016/j.cognition.2010.12.011

Deak, G. O., Yen, L., & Pettit, J. (2001). By any other name: when will preschoolers produce several labels for a referent?. *Journal of Child Language*, *28*(3), 787-804. https://doi.org/10.1017/S0305000901004858

Diesendruck, G., & Markson, L. (2001). Children's avoidance of lexical overlap: A pragmatic account. *Developmental Psychology, 37*, 630–641. doi: 10.1037/0012-1649.37.5.630

Dunn, D. M., & Dunn, L. M. (2007). *PPVT-4: Peabody picture vocabulary test.* Pearson Assessments.

Estis, J. M., & Beverly, B. L. (2015). Children with SLI exhibit delays resolving ambiguous reference. *Journal of Child Language*, *42*, 180–195. https://doi.org/10.1017/S0305000914000038

Evey, J. A., & Merriman, W. E. (1998). The prevalence and the weakness of an early name mapping preference. *Journal of Child Language*, *25*(1), 121-147.

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American Psychologist*, *34*(10), 906. https://psycnet.apa.org/doi/10.1037/0003-066X.34.10.906

Gathercole, V. C. (1989). Contrast: A semantic constraint? *Journal of Child Language, 16*, 685–702. doi: 10.1017/S0305000900010795

Gliga, T., & Csibra, G. (2009). One-year-old infants appreciate the referential nature of deictic gestures and words. *Psychological Science, 20(3),* 347-353. https://doi.org/10.1111/j.1467-9280.2009.02295.x

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. doi: https://psycnet.apa.org/doi/10.1037/0012-1649.28.1.99

Gollek, C., & Doherty, M. J. (2016). Metacognitive developments in word learning: Mutual exclusivity and theory of mind. *Journal of experimental child psychology*, *148*, 51-69. https://doi.org/10.1016/j.jecp.2016.03.007

Graham, S. A., Nilsen, E. S., Collins, S., & Olineck, K. (2010). The role of gaze direction and mutual exclusivity in guiding 24-month-olds' word mappings. *British Journal of Developmental Psychology, 28*, 449–465. https://doi.org/10.1348/026151009X424565

Grassmann, S., & Tomasello, M. (2010). Young children follow pointing over words in interpreting acts of reference. *Developmental Science*, *13*(1), 252-263. https://doi.org/10.1111/j.1467-7687.2009.00871.x

Halberda, J. (2003). The development of a word-learning strategy. *Cognition, 87,* B23–B34. doi: 10.1016/S0010-0277(02)00186-5

Halberda, J. (2006). Is this a dax which I see before me? Use of the logical arguments disjunctive syllogism supports word-learning in children and adults. *Cognitive Psychology, 53*, 310–344. doi: 10.1016/j.cogpsych.2006.04.003

Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children*. Paul H Brookes Publishing.

Hartin, T. L., Stevenson, C. M., & Merriman, W. E. (2016). Preexposure to objects that contrast in familiarity improves young children's lexical knowledge judgment. *Language Learning and Development*, *12*(3), 311-327. https://doi.org/10.1080/15475441.2015.1077123

Henning, K. J. & Merriman, W. E., (2019). The disambiguation prediction effect. *Journal of Cognition and Development*. Advance online publication. doi.org/10.1080/15248372.2019.1578224

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

Horst, J. S., Scott, E. J., & Pollard, J. A. (2010). The role of competition in word learning via referent selection. *Developmental Science*, *13*(5), 706-713. https://doi.org/10.1111/j.1467-7687.2009.00926.x

Horst, J. S., Samuelson, L. K., Kucker, S. C., & McMurray, B. (2011). What's new? Children prefer novelty in referent selection. *Cognition*, *118*(2), 234-244. https://doi.org/10.1016/j.cognition.2010.10.015

Houston-Price, C., Caloghiris, Z., & Raviglione, E. (2010). Language experience shapes the development of the mutual exclusivity bias. *Infancy, 15*, 125–150. https://doi.org/10.1111/j.1532-7078.2009.00009

Hughes, C., & Dunn, J. (1998). Understanding mind and emotion: longitudinal associations with mental-state talk between young friends. *Developmental psychology*, *34*(5), 1026. doi: http://dx.doi.org/10.1037/0012-1649.34.5.1026

Jameson, K. A., Narens, L., Goldfarb, K., & Nelson, T. O. (1990). The influence of near-threshold priming on metamemory and recall. *Acta Psychologica, 73(1)*, 55-68. https://doi.org/10.1016/0001-6918(90)90058-N

Jarvis, L. H., Merriman, W. E., Barnett, M., Hanba, J., & Van Haitsma, K. S. (2004). Input that contradicts young children's strategy for mapping novel words affects their phonological and semantic interpretation of other novel words. *Journal of Speech, Language, and Hearing Research*. https://doi.org/10.1044/1092-4388(2004/032)

Jaswal, V.K., & Hansen, M.B. (2006). Learning words: children disregard some pragmatic information that conflicts with mutual exclusivity. *Developmental Science, 9*, 158–165. https://doi.org/10.1111/j.1467-7687.2006.00475.x

Johnson, C. N., & Maratsos, M. P. (1977). Early comprehension of mental verbs: Think and know. *Child Development*, 1743-1747. 10.2307/1128549

Kalashnikova, M., Mattock, K., & Monaghan, P. (2014). Disambiguation of novel labels and referential facts: A developmental perspective. *First Language*, *34*(2), 125-135. https://doi.org/10.1177%2F0142723714525946

Karmiloff-Smith, A. (1986). From meta-processes to conscious access: Evidence from children's metalinguistic and repair data. *Cognition*, *23*(2), 95-147. https://doi.org/10.1016/0010-0277(86)90040-5

Lederberg, A. R., & Spencer, P. E. (2008). Word-learning abilities in deaf and hard-of-hearing preschoolers: Effect of lexicon size and language modality. *Journal of Deaf Studies and Deaf Education*, *14*(1), 44-62. https://doi.org/10.1093/deafed/enn021

Lewis, M., & Frank, M. C. (2015). A meta-analytic approach to understanding the disambiguation effect. Proceedings from the 2015 Biennial Meeting of the Society for Research in Child Development. Philadelphia, PA.

Lewis, M., Cristiano, V., Lake, B. M., Kwan, T., & Frank, M. C. (2019). The role of developmental change and linguistic experience in the mutual exclusivity effect. *PsyArXiv Preprints.* https://doi.org/10.31234/osf.io/wsx3a

Lipowski, S. L., Merriman, W. E., & Dunlosky, J. (2013). Preschoolers can make highly accurate judgments of learning. *Developmental Psychology*, *49*(8), 1505. https://psycnet.apa.org/doi/10.1037/a0030614

Low, J., & Perner, J. (2012). Implicit and explicit theory of mind: State of the art. *British Journal of Developmental Psychology*, *30*(1), 1-13. doi: https://doi.org/10.1111/j.2044-835X.2011.02074.x

Marazita, J. M., & Merriman, W. E. (2004). Young children's judgment of whether they know names for objects: The metalinguistic ability it reflects and the processes it involves. *Journal of Memory and Language, 51*, 458–472. doi: 10.1016/j.jml.2004.06.008

Markman, E. M. (1984). The acquisition and hierarchical organization of categories by children. In C. Sophian (Ed.), *Origins of cognitive skills: The 18th annual Carnegie Symposium on Cognition* (pp. 376-406). Hillsdale, NJ: Erlbaum.

Markman, E. M., & Wachtel, G. F. (1989). Children's use of mutual exclusivity to constrain the meaning of words. *Cognitive Psychology, 20*, 121–157. doi: 10.1016/0010-0285(88)90017-5

Markman, E. M., Wasow, J. L., & Hansen, M. B. (2003). Use of mutual exclusivity assumption by young word learners. *Cognitive Psychology, 47*, 241–275. doi: 10.1016/S0010-0285(03)00034-3

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior research methods*, *44*(2), 314-324.

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*, 831–877. doi: 10.1037/a0029872

Merriman, W. E. (1986). Some reasons for the occurrence and eventual correction of children's naming errors. *Child Development, 57*, 942-952. doi:10.2307/1130370

Merriman, W. E. (1999). Competition, attention, and young children's lexical processing. In B. MacWhinney (Ed.), *The emergence of language* (pp.331–358). Hillsdale, NJ: Erlbaum.

Merriman, W. E., & Bowman, L. L. (1989). The mutual exclusivity bias in children's word learning. *Monographs of the Society for Research in Child Development, 54*(3–4, Serial No. 220) doi: 10.2307/1166130.

Merriman, W. E., & Schuster, J. M. (1991). Young children's disambiguation of object name reference. *Child Development, 62*, 1288–1301. doi: 10.1111/j.1467-8624.1991.tb01606.x

Merriman, W. E., Marazita, J., & Jarvis, L. H. (1993). Four-Year-Olds' Disambiguation of Action and Object Word Reference. *Journal of Experimental Child Psychology*, *56*(3), 412-430. https://doi.org/10.1006/jecp.1993.1042

Merriman, W. E., & Marazita, J. M. (1995). The effect of hearing similar-sounding words on young 2-year-olds' disambiguation of novel noun reference. *Developmental Psychology, 31*, 973–984. doi: 10.1037/0012-1649.31.6.973

Merriman, W. E., Marazita, J., & Jarvis, L. (1995). Children's disposition to map new words onto new referents. *Beyond names for things: Young children's acquisition of verbs*, 147-183.

Merriman, W. E., & Lipko, A. R. (2008). A dual criterion account of the development of linguistic judgment in early childhood. *Journal of Memory and Language, 58,* 1012–1031. doi: 10.1016/j.jml.2007.10.003

Merriman, W. E., Lipko, A. R., & Evey, J. A. (2008). How young children judge whether a word is one they know: A dual criterion account. *Journal of Experimental Child Psychology, 101*, 83-98. doi: 10.1016/j.jecp.2008.06.001

Mervis, C. B., & Bertrand, J. (1994). Acquisition of the novel name–nameless category (N3C) principle. *Child Development, 65*, 1646–1662. doi: 10.1111/j.1467-8624.1994.tb00840.x

Mervis, C., & Bertrand, J. (1995a). Early lexical acquisition and the vocabulary spurt: A response to Goldfield & Reznick. *Journal of Child Language, 22*, 461-468. doi:10.1017/S0305000900009880

Mervis, C. B., & Bertrand, J. (1995b). Acquisition of the novel name-nameless category (N3C) principle by young children who have down syndrome. *American Journal on Mental Retardation, 100,* 231-43.

Moore, C., Bryant, D., & Furrow, D. (1989). Mental terms and the development of certainty. *Child Development*, 167-171. 10.2307/1131082

Moore, C., Furrow, D., Chiasson, L., & Patriquin, M. (1994). Developmental relationships between production and comprehension of mental terms. First language, 14(42-43), 001-17. doi: https://doi.org/10.1177/014272379401404201

O'Neill, D. K., Astington, J. W., & Flavell, J. H. (1992). Young children's understanding of the role that sensory experiences play in knowledge acquisition. *Child Development*, *63*(2), 474-490. doi: https://doi.org/10.1111/j.1467-8624.1992.tb01641.x

Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science*, *308*(5719), 255-258. doi: 10.1126/science.1107621

Pace, A., Luo, R., Hirsh-Pasek, K., & Golinkoff, R. M. (2017). Identifying pathways between socioeconomic status and language development. *Annual Review of Linguistics*, *3*, 285-308. doi: https://doi.org/10.1146/annurev-linguistics-011516-034226

Preissler, M., & Carey, S. (2005). The role of inferences about referential intent in word learning: Evidence from autism. *Cognition, 97*, B13–B23. https://doi.org/10.1016/j.cognition.2005.01.008

Regier, T. (2005). The emergence of words: Attentional learning in form and meaning. *Cognitive Science, 29*, 115–135. doi: 10.1207/s15516709cog0000_31

Roby, E., & Scott, R. M. (2018). The relationship between parental mental-state language and 2.5-year-olds' performance on a nontraditional false-belief task. *Cognition*, *180*, 10-23. doi: https://doi.org/10.1016/j.cognition.2018.06.017

Romski, M. A., Sevcik, R. A., Robinson, B. F., Mervis, C. B., & Bertrand, J. (1996). Mapping the meanings of novel visual symbols by youth with moderate or severe mental retardation. *American Journal on Mental Retardation, 100*, 391–402.

Rosenthal, R. (1990). How are we doing in soft psychology?. *American Psychologist*, *45*(6), 775. doi: https://psycnet.apa.org/doi/10.1037/0003-066X.45.6.775

Rosenthal, R. (1991). Meta-analytic procedures for social research. Newbury Park: CA: Sage.

Scofield, J., & Behrend, D. A. (2007). Two-year-olds differentially disambiguate novel words and facts. *Journal of Child Language*, *34*(4), 875-889. https://doi.org/10.1017/S0305000907008100

Scofield, J., Merriman, W. E., & Wall, J. L. (2018). The effect of a tactile-to-visual shift on young children's tendency to map novel labels onto novel objects. *Journal of experimental child psychology*, *172*, 1-12. https://doi.org/10.1016/j.jecp.2018.03.003

Scott, R. M., Richman, J. C., & Baillargeon, R. (2015). Infants understand deceptive intentions to implant false beliefs about identity: New evidence for early mentalistic reasoning. *Cognitive psychology*, *82*, 32-56. doi: https://doi.org/10.1016/j.cogpsych.2015.08.003

Slocum, J. Y., & Merriman, W. E. (2018). The metacognitive disambiguation effect. *Journal of Cognition and Development*, *19*(1), 87-106. https://doi.org/10.1080/15248372.2017.1415901

Smith, C. L., & Tager-Flusberg, H. (1982). Metalinguistic awareness and language development. *Journal of Experimental Child Psychology, 34*, 449–468. doi: 10.1016/0022-0965(82)90071-6

Sodian, B., & Wimmer, H. (1987). Children's understanding of inference as a source of knowledge. *Child development*, 424-433. DOI: 10.2307/1130519

Spiegel, C., & Halberda, J. (2011). Rapid fast-mapping abilities in 2-year-olds. *Journal of experimental child psychology*, *109*(1), 132-140. https://doi.org/10.1016/j.jecp.2010.10.013

Surian, L., Caldi, S., & Sperber, D. (2007). Attribution of beliefs by 13-month-old infants. *Psychological science*, *18*(7), 580-586. doi: https://doi.org/10.1111%2Fj.1467-9280.2007.01943.x

Treisman, A. (1985). Preattentive processing in vision. *Computer vision, graphics, and image processing*, *31*(2), 156-177. https://doi.org/10.1016/S0734-189X(85)80004-9

Venetsanou, F., Kambas, A., Aggeloussis, N., Fatouros, I., & Taxildaris, K. (2009). Motor assessment of preschool aged children: A preliminary investigation of the validity of the Bruininks–Oseretsky test of motor proficiency. *Human Movement Science*, *28*(4), 543-550. doi: https://doi.org/10.1016/j.humov.2009.03.002

Valdez-Menchaca, M. C., & Whitehurst, G. J. (1992). Accelerating language development through picture book reading: a systematic extension to Mexican day care. *Developmental psychology*, *28*(6), 1106. doi: https://psycnet.apa.org/doi/10.1037/0012-1649.28.6.1106

Wall, J. L., Merriman, W. E., Scofield, J. (2015). Young children's disambiguation across the senses. *Cognitive Development, 35*, 163-177. doi: 10.1016/j.cogdev.2015.06.001

Waxman, S. R., & Senghas, A. (1992). Relations among word meanings in early lexical development. *Developmental Psychology, 28*, 862-873. doi:10.1037/0012-1649.28.5.862

Wilkinson, K., & Green, G. (1998). Implications of fast mapping for vocabulary expansion in individuals with mental retardation. *Augmentative and Alternative Communication, 14:3*, 162-170. doi: 10.1080/07434619812331278336

Wolfe, J. M. (1994). Guided search 2.0 a revised model of visual search. *Psychonomic bulletin & review*, *1*(2), 202-238.

Wooden, E. W. (2006). The MacArthur-Bates Communicative Development Inventory as an assessment tool for low-income, African American children.

Yurovsky, D., Smith, L. B., & Yu, C. (2013). Statistical word learning at scale: The baby's view is better. *Developmental science, 16(6)*, 959-966. https://doi.org/10.1111/desc.12036

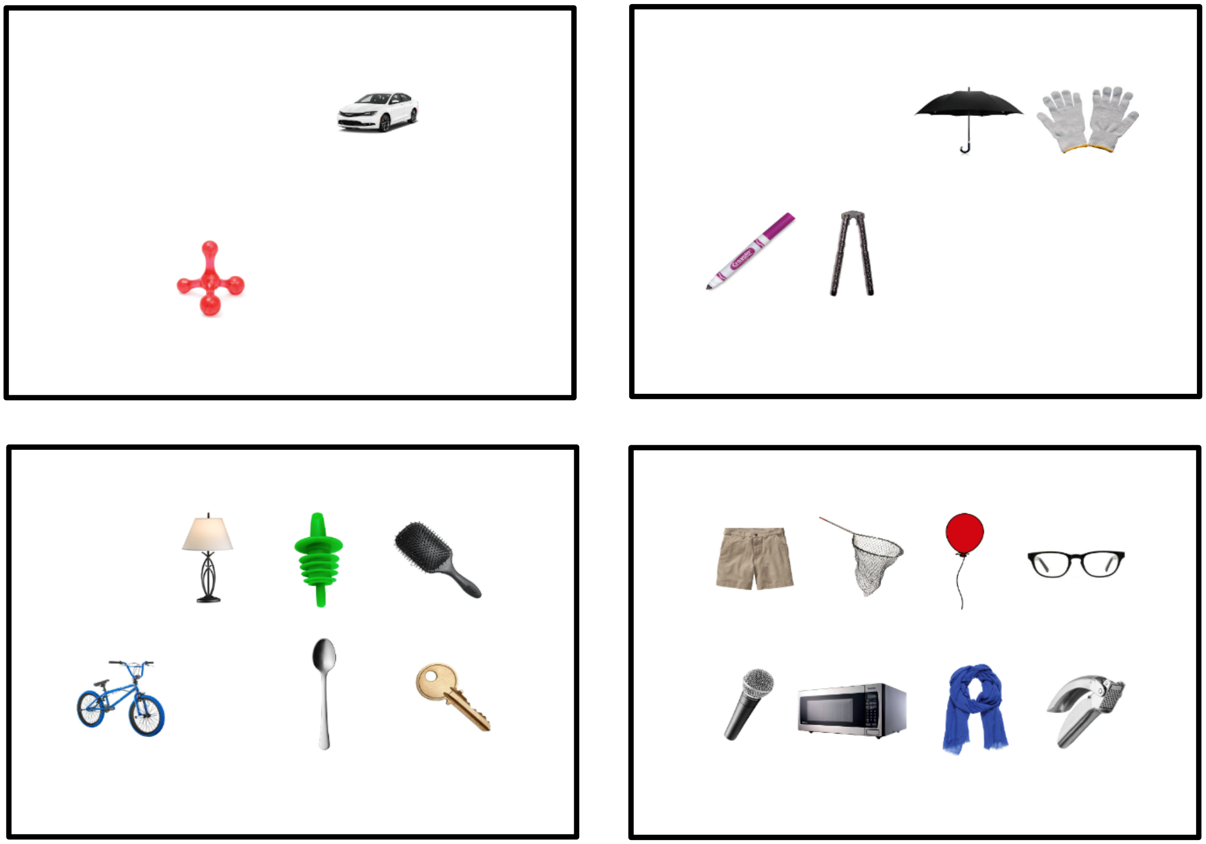


Figure 1. *Examples of slides for 2-, 4-, 6-, and 8-choice problems in Experiment 1.*

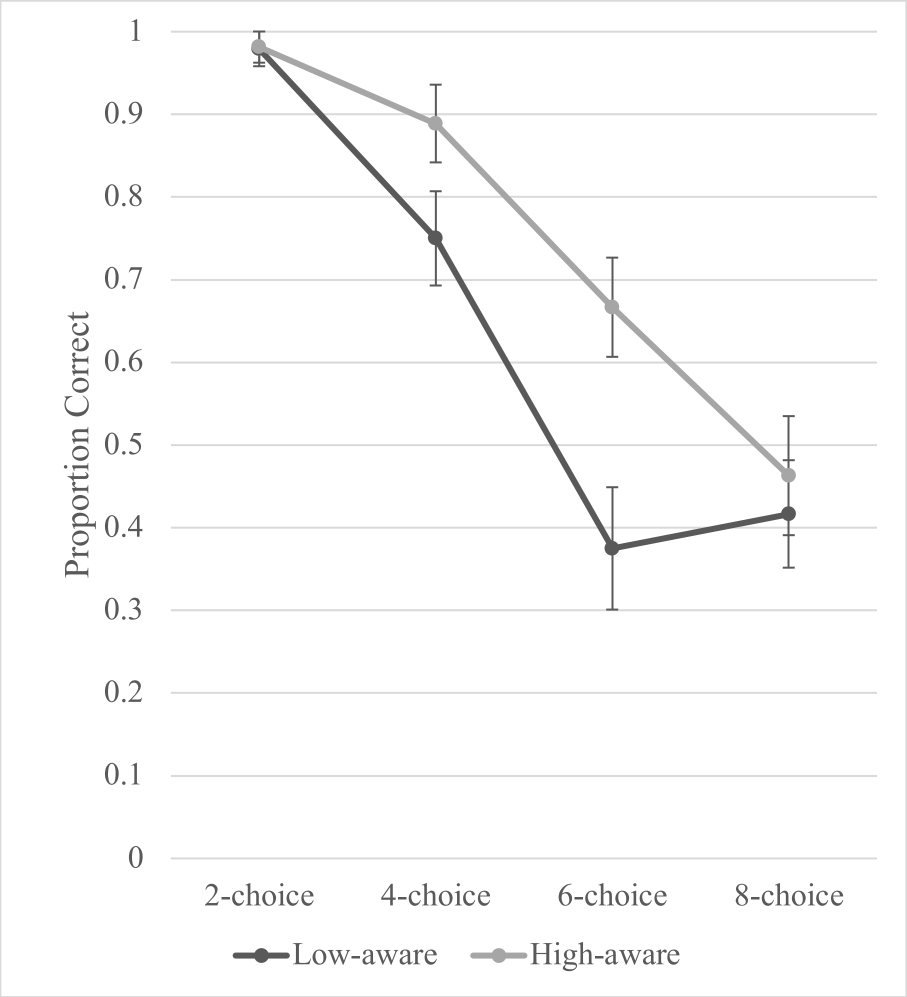


Figure 2. Novel label mapping accuracy by problem type for children with high and low awareness of lexical knowledge. Error bars represent one standard error above and below.

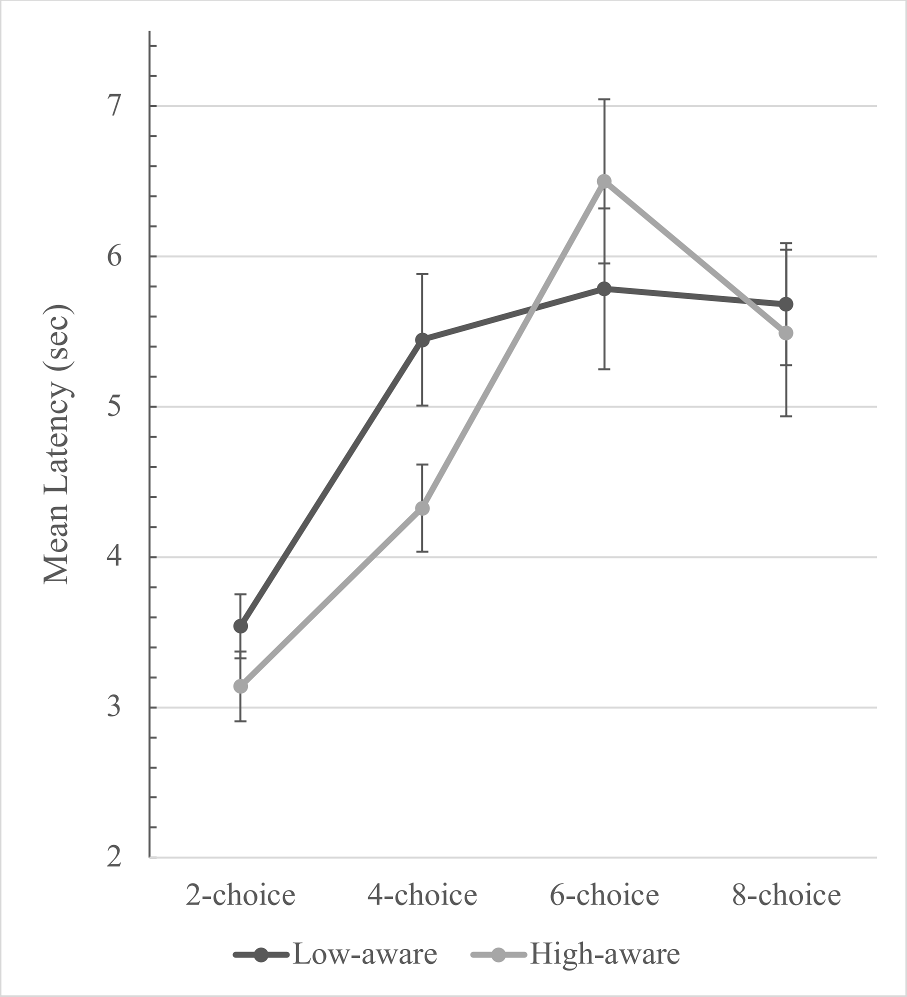
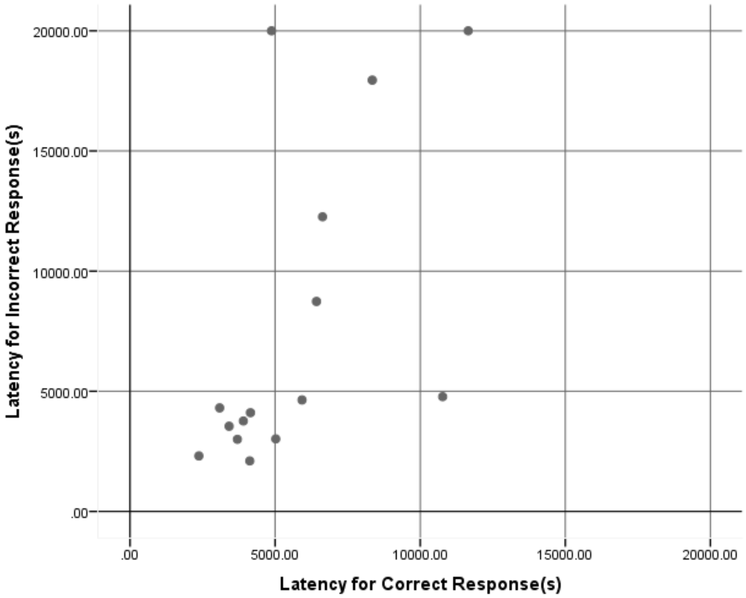
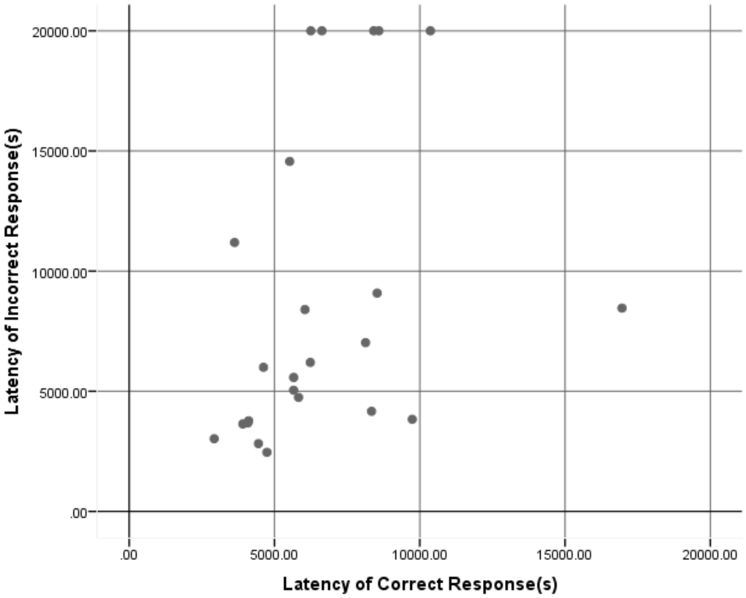


Figure 3. Novel label mapping latency by problem type for children with high and low word knowledge judgment performance. Error bars represent one standard error above and below.

Four-Choice Problems



Six-Choice Problems



Eight-Choice Problems

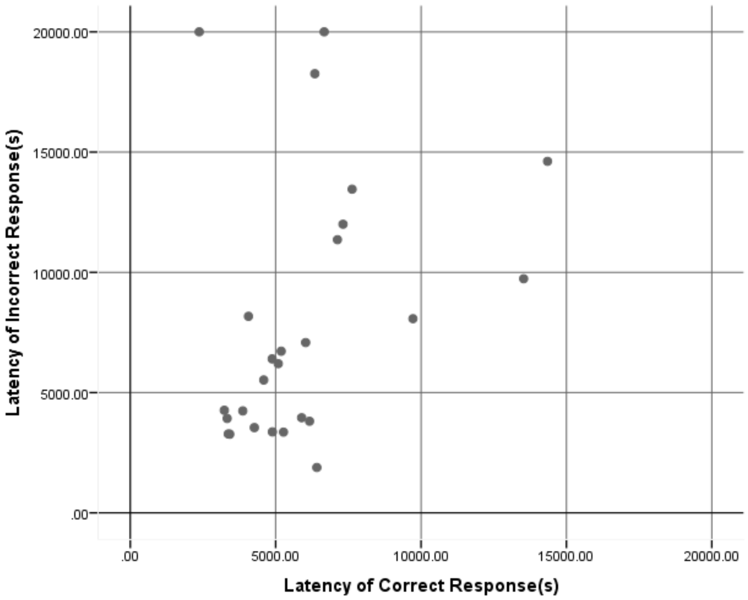


Figure 4. *Latencies of correct and incorrect responses on novel label trials in Experiment 1.*

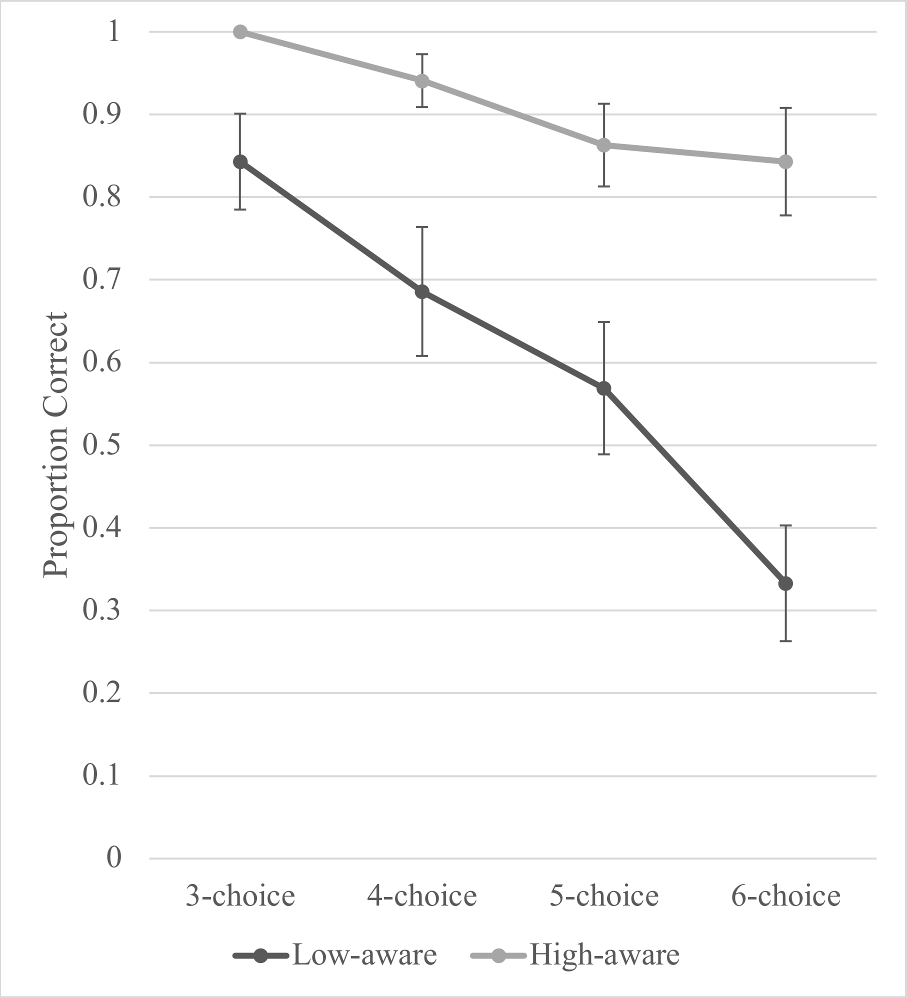


Figure 5. Novel name mapping accuracy by problem type for children with high and low awareness of lexical knowledge in Experiment 2. Error bars represent one standard error above and below.

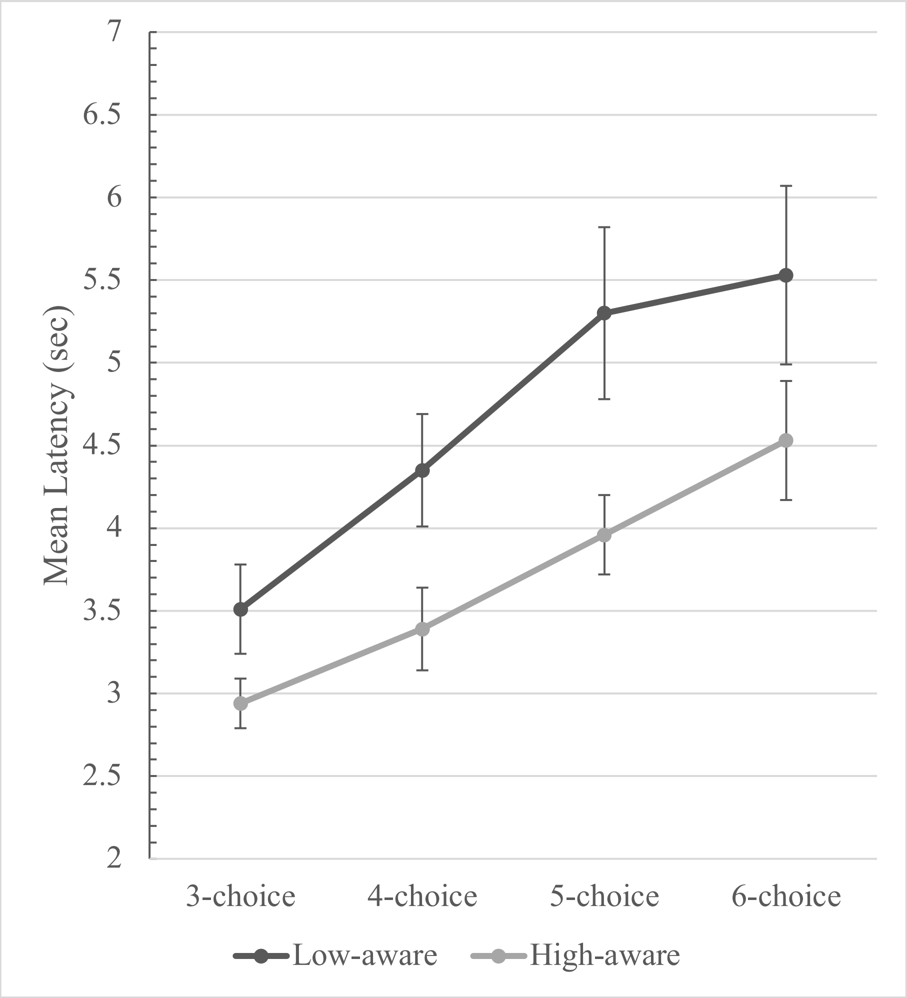
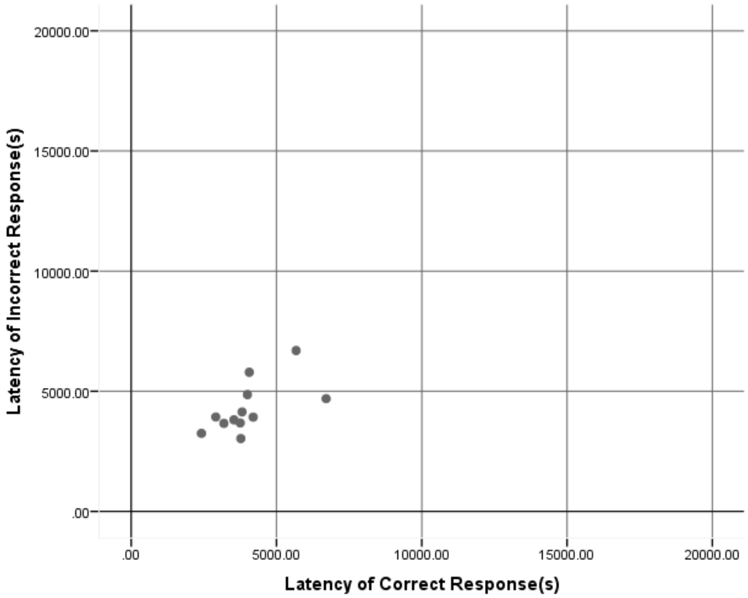
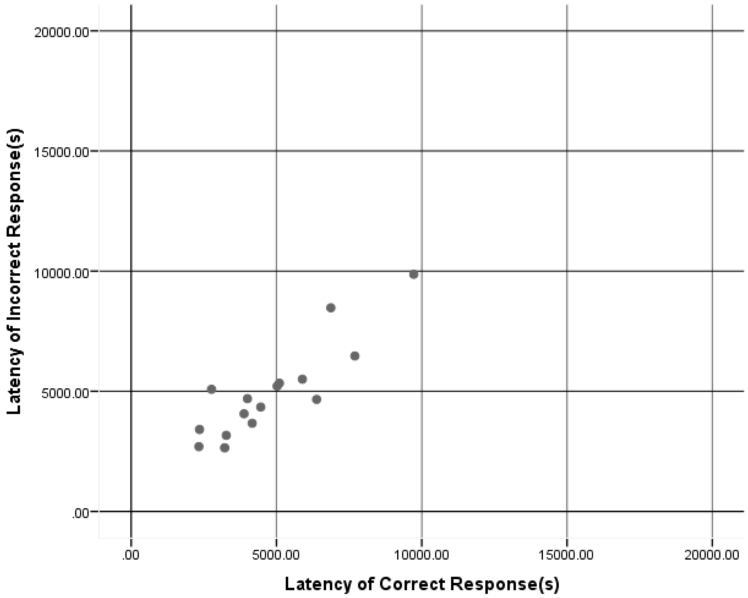


Figure 6. *Novel name mapping latency by problem type for children with high and low object nameability judgment performance in Experiment 2. Error bars represent one standard error above and below.*

Four-Choice Problems



Five-Choice Problems



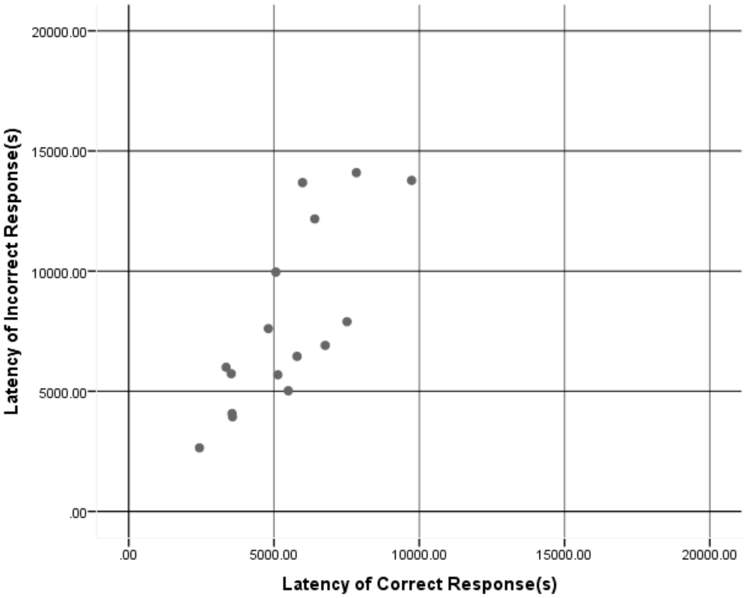


Figure 7. Latencies of correct and incorrect responses on novel label trials in Experiment 2.

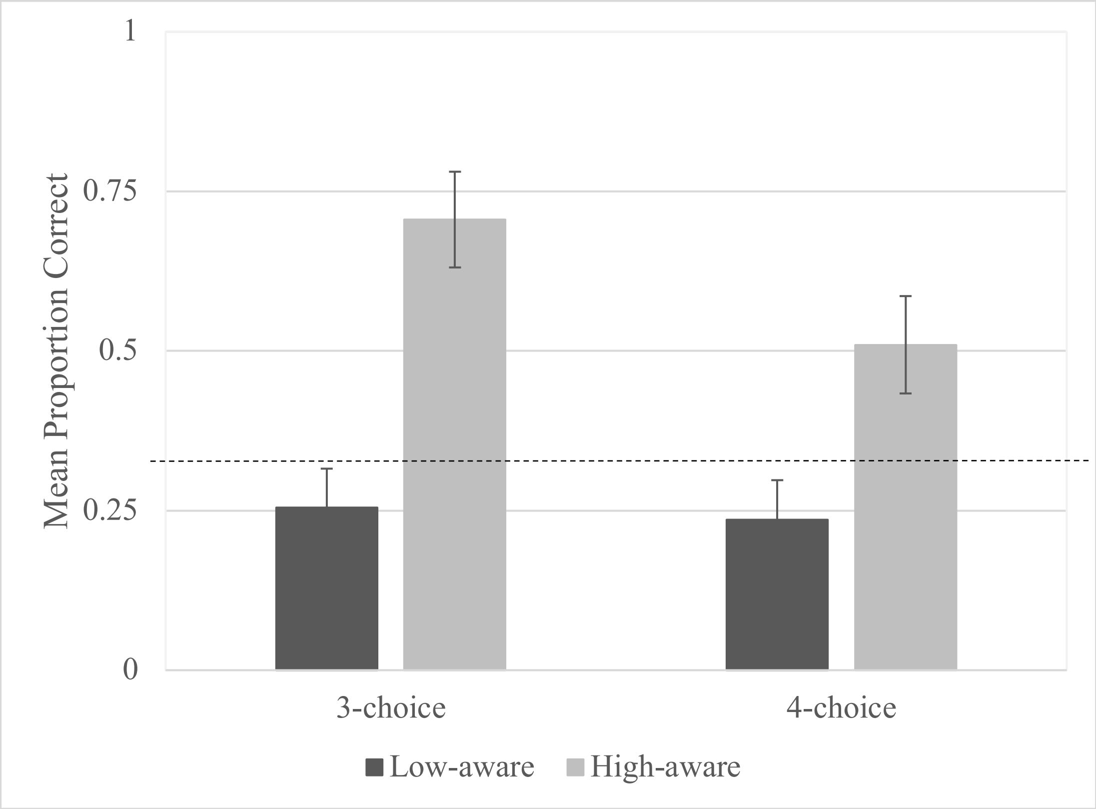


Figure 8. *Proportion of children's correct choices on the retention trials. The dotted line represents chance (.33). Error bars represent one standard error above and below.*

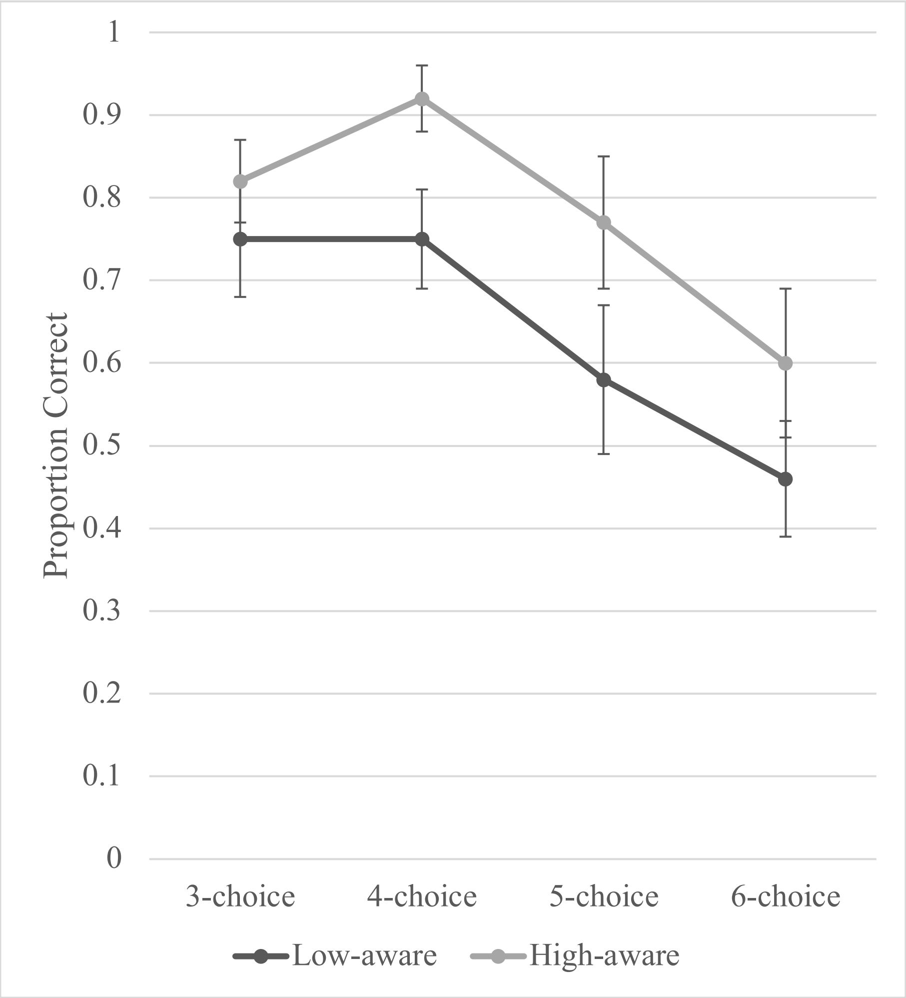


Figure 9. Novel name mapping accuracy by problem type for children with high and low awareness of lexical knowledge in Experiment 3. Error bars represent one standard error above and below.

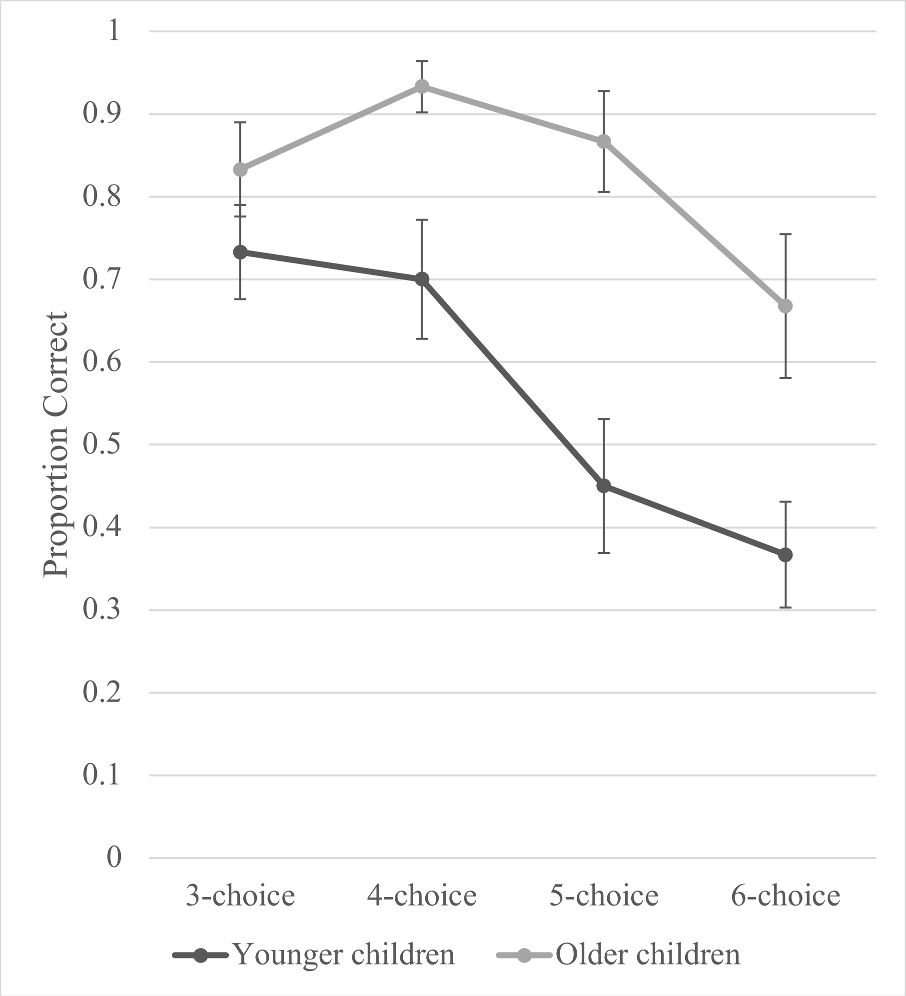


Figure 10. Novel name mapping accuracy by problem type for children who were older and younger than the median age in Experiment 3. Error bars represent one standard error above and below.

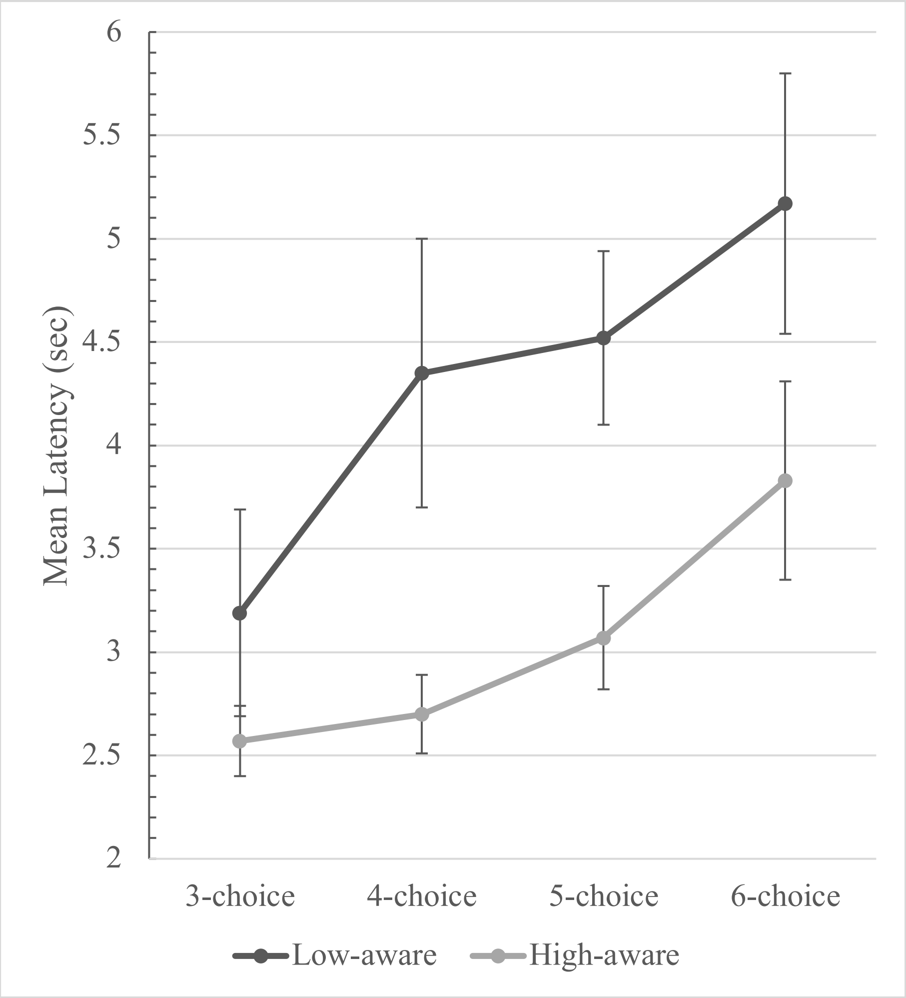
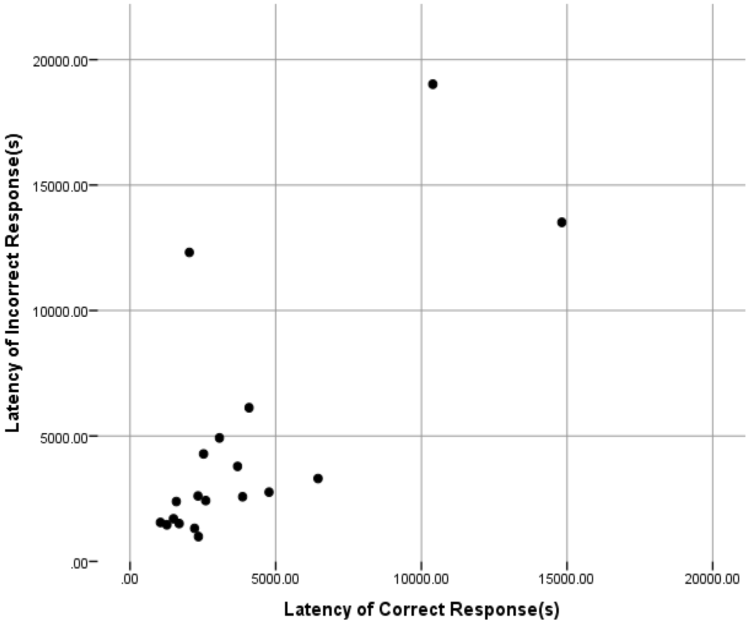
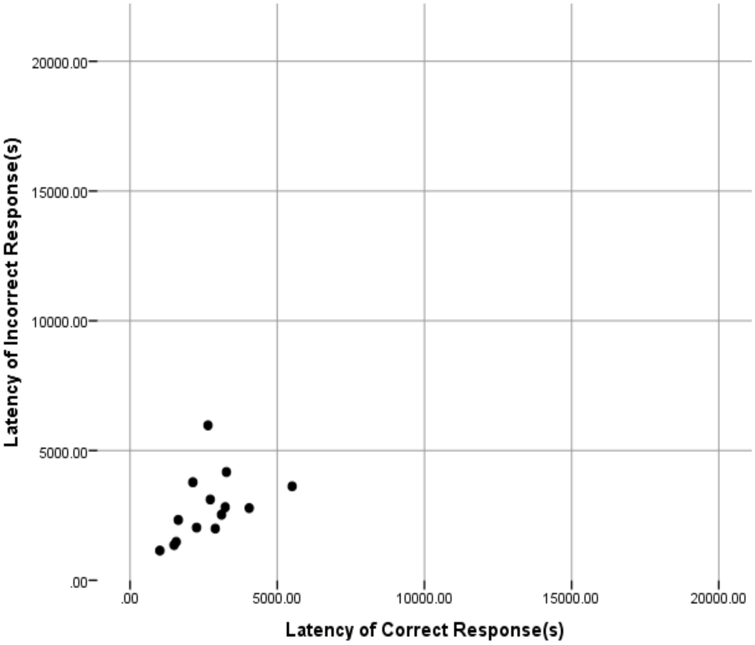


Figure 11. Novel name mapping latency by problem type for children with high and low awareness of lexical knowledge in Experiment 3. Error bars represent one standard error above and below.

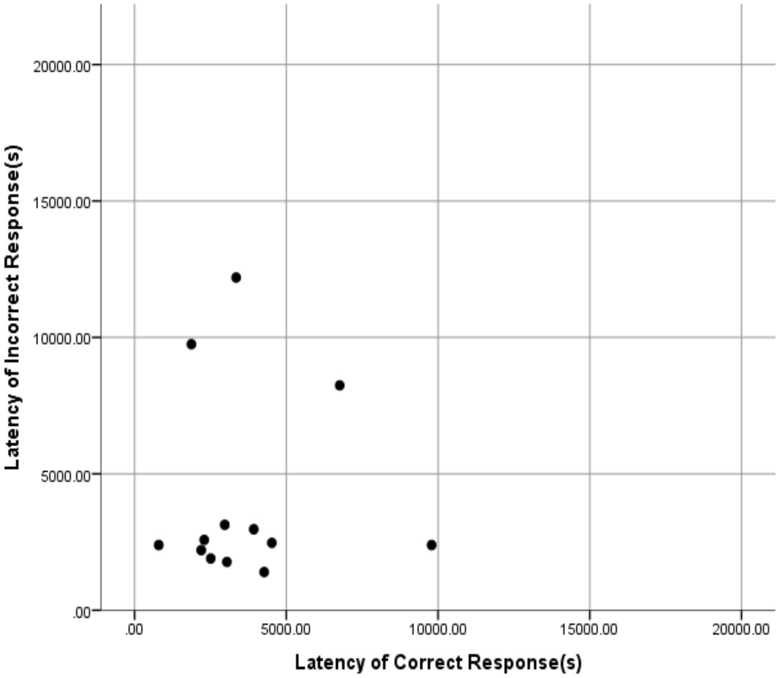
Three-Choice Problems



Four-Choice Problems



Five-Choice Problems



Six-Choice Problems

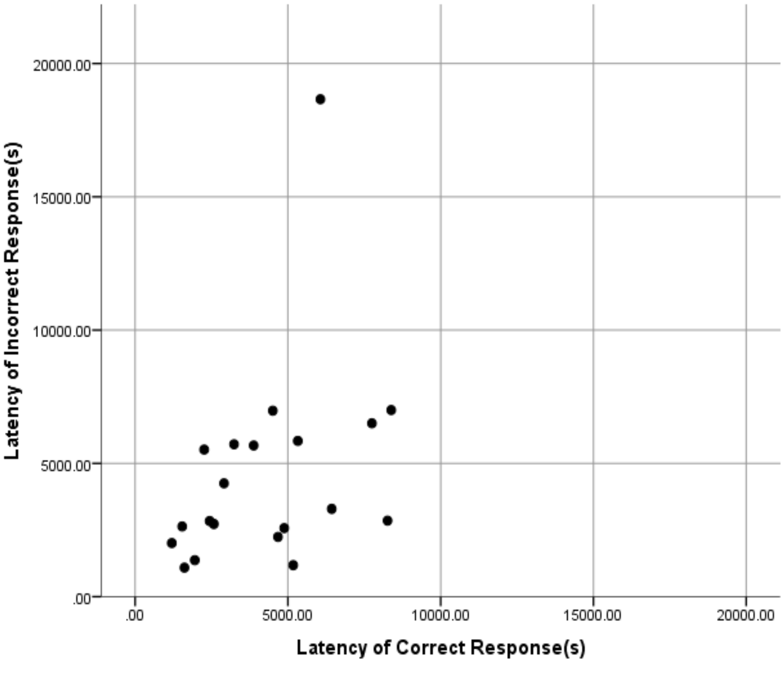


Figure 12. Latencies of correct and incorrect responses on novel label trials in Experiment 3.

	Disambiguation				Word Knowledge			Obj Nameability		
Choices:	2	4	6	8	Fam	Unfam	Avg	Fam	Unfam	Avg
	.98	.82	.53	.44	.99	.67	.83	.96	.73	.85
	(.08)	(.22)	(.31)	(.28)	(.05)	(.36)	(.18)	(.09)	(.29)	(.15)

Table 1. Mean proportion correct (SD) in the tasks of Experiment 1.

MEASURE	2	3	4
1. Age	.10	.31[a]	.23
2. Disambiguation		.39*	.55**
3. Word Knowledge			.55**
4. Object Nameability			

Notes- [a] *one-tailed* $p < .05$ * *two-tailed* $p < .05$ ** *two-tailed* $p < .01$

Table 2. Task intercorrelations and correlations with age in Experiment 1.

	CHOICES			
LABEL	2	4	6	8
Familiar				
M	2.17	2.48	2.68	3.21
SD	*.49*	*.39*	*.63*	*.77*
Novel				
M	3.33	4.85	6.21	5.58
SD	*.93*	*1.58*	*2.14*	*1.84*
n	*34*	*34*	*30*	*29*

Table 3. Latency of correct responses (in sec) in the referent selection task of Experiment 1.

	CHOICES					
PREDICTOR	2	4	6	8	Overall Mean	Linear Trend
1. Age	-.31[a]	-.45**	-.38*	-.37*	- .52**	-.21
2. Word Knowledge	.03	-.05	-.14	-.17	-.14	-.24
3. Object Nameability	-.15	-.25	-.35*	-.36*	-.41**	-.35*

*Notes- [a] one-tailed p < .05 * two-tailed p < .05 ** two-tailed p < .01 df = 32*

Table 4. Correlations between predictors and mean latency to map familiar labels in Experiment 1.

	CHOICES					
PREDICTOR	2	4	6	8	Overall Mean	Linear Trend
1. Age	-.31[a]	-.32[a]	-.27	-.39*	- .47*	-.35[a]
2. Word Knowledge	-.29[a]	-.33[a]	+.31[a]	-.11	-.08	+.18
3. Object Nameability	-.21	-.22	-.07	-.17	-.33	-.14

Notes- [a] *one-tailed $p < .05$ * two-tailed $p < .05$ ** two-tailed $p < .01$*

Correlations are for pairwise comparisons: df = 32 for 2- and 4-choice; 28 for 6-choice; 29 for 8-choice; 24 for overall mean and linear trend.

Table 5. Correlations between predictors and mean latencies of correct response on novel label trials in Experiment 1.

Disambiguation				Word Knowledge			Obj Nameability		
Choices: 3	4	5	6	Fam	Unfam	Avg	Fam	Unfam	Avg
.92 (.18)	.81 (.27)	.72 (.31)	.59 (.38)	.95 (.18)	.65 (.42)	.79 (.22)	.96 (.04)	.76 (.33)	.88 (.17)

Table 6. Mean proportion correct (SD) in the tasks of Experiment 2.

MEASURE	2	3	4	5	6
1. Age	.72**	.52**	.56**	.69**	.35*
2. Disambig Accuracy		.72**	.62**	.79**	.23
3. Disambig Retention			.53**	.71**	.31[a]
4. Word Knowledge				.63**	.21
5. Object Nameability					.16
6. Vocabulary Size					

Notes- [a] *one-tailed p < .05 * two-tailed p < .05 ** two-tailed p < .01*

Correlations are for listwise comparisons (df = 30)

Table 7. Task intercorrelations and correlations with age in Experiment 2.

	CHOICES			
LABEL	3	4	5	6
Familiar				
M	2.39	2.78	2.91	2.92
SD	.44	.61	.59	.69
Novel				
M	3.36	3.94	4.69	5.03
SD	.87	1.21	1.56	1.63
n	34	33	31	29

Table 8. Latency of correct responses (in sec) in the referent selection task of Experiment 2.

	CHOICES					
PREDICTOR	3	4	5	6	Overall Mean	Linear Trend
1. Age	-.59**	-.47**	-.40*	-.63**	- .65**	-.27
2. Word Knowledge	-.08	-.28	-.01	-.21	-.19	-.09
3. Object Nameability	-.34	-.27	-.08	-.46*	-.36*	-.21
4. Vocabulary Size	-.08	-.29	-.09	-.13	-.19	-.03

Notes- ^{a} *one-tailed p < .05 * two-tailed p < .05 ** two-tailed p < .01*

Correlations are for listwise comparisons (df = 30)

Table 9. Correlations between predictors and mean latency to map familiar names in Experiment 2.

	CHOICES					
PREDICTOR	3	4	5	6	Overall Mean	Linear Trend
1. Age	-.52**	-.28	-.46*	-.32[a]	-.38[a]	-.18
2. Word Knowledge	-.55**	-.39*	-.55**	-.31	-.49*	-.19
3. Object Nameability	-.59**	-.46**	-.56**	-.60**	-.64**	-.45*
4. Vocabulary Size	-.23	-.32[a]	-.39*	-.56**	-.46*	-.49*

Notes- [a] *one-tailed p < .05 * two-tailed p < .05 ** two-tailed p < .01*

Correlations are for pairwise comparisons: df = 30 for 3-choice; 29 for 4-choice; 27 for 5-choice; 25 for 6-choice; 23 for overall mean and linear trend.

Table 10. Correlations between predictors and mean latencies of correct response on novel label trials in Experiment 2.

	Disambiguation				Word Knowledge			Obj Nameability		
Choices:	3	4	5	6	Fam	Unfam	Avg	Fam	Unfam	Avg
	.79	.83	.65	.51	.92	.62	.77	.95	.75	.85
	(.25)	(.27)	(.38)	(.36)	(.19)	(.43)	(.20)	(.11)	(.34)	(.17)

Table 11. Mean proportion correct (SD) in the tasks of Experiment 3.

MEASURE	2	3	4	5
1. Age	.66**	.46**	.36*	.66**
2. Disambig Accuracy		.44*	.32[a]	.55**
3. Word Knowledge			.60**	.52**
4. Object Nameability				.41*
5. Vocabulary Size				

*Notes- [a] one-tailed p < .05 * two-tailed p < .05 ** two-tailed p < .01*

Correlations were computed listwise (df = 36)

Table 12. Task intercorrelations and correlations with age in Experiment 3.

	CHOICES			
LABEL	3	4	5	6
Familiar				
M	1.69	1.73	1.81	1.96
SD	.66	.64	.63	.72
Novel				
M	*3.05*	*3.45*	*3.72*	*4.50*
SD	*1.68*	2.02	1.61	2.06
n	39	38	33	32

Table 13. Latency of correct responses (in sec) in the referent selection task of Experiment 3.

	CHOICES					
PREDICTOR	3	4	5	6	Overall Mean	Linear Trend
1. Age	.09	-.19	-.22	-.28	-.19	-.29[a]
2. Word Knowledge	-.13	-.26	-.42**	-.27	-.35*	-.14
3. Object Nameability	-.07	-.14	-.22	-.26	-.22	-.16
4. Vocabulary Size	.04	-.22	-.31[a]	-.26	-.24	-.25

*Notes- [a] one-tailed p < .05 * two-tailed p < .05 ** two-tailed p < .01*

Correlations were computed listwise (df = 36)

Table 14. Correlations between predictors and log-transformed mean latency to map familiar names in Experiment 3.

	CHOICES					
PREDICTOR	3	4	5	6	Overall Mean	Linear Trend
1. Age	.16	.31[a]	.32[a]	-.02	.16	-.07
2. Word Knowledge	.16	-.02	-.16	-.13	-.19	-.31
3. Object Nameability	.23	.00	-.10	-.10	-.20	-.35[a]
4. Vocabulary Size	.22	.27	.17	-.07	+.15	-.44*

Notes- [a] *one-tailed* $p < .05$ * *two-tailed* $p < .05$ ** *two-tailed* $p < .01$

Correlations are for pairwise comparisons: $df = 35$ for 3- and 4-choice; 30 for 5-choice; 29 for 6-choice; 24 for overall mean and linear trend.

Table 15. Correlations between predictors and log-transformed mean latencies of correct response on novel label trials in Experiment 3.

	CHOICES					
PREDICTOR	3	4	5	6	Overall Mean	Linear Trend
1. Age	.32	.10	.45*	-.05	.28	-.09
2. Word Knowledge	.22	.02	-.10	-.04	.13	-.02
3. Object Nameability	.04	.07	-.19	-.01	.00	-.22
4. Vocabulary Size	.17	.16	.42*	-.08	.24	-.07

Notes- ^{a} *one-tailed* $p < .05$ * *two-tailed* $p < .05$ ** *two-tailed* $p < .01$

Correlations were computed listwise (df = 35)

Table 16. Correlations between predictors and number of foils checked on novel label trials in Experiment 3.

PREDICTOR	CHOICES 3	4	5	6	Overall Mean	Linear Trend
1. Age	.29^{a}	.37*	.12	-.11	.30^{a}	-.24
2. Word Knowledge	.19	.11	.07	.07	.20	-.03
3. Object Nameability	.21	.11	-.15	-.16	.05	-.18
4. Vocabulary Size	.30^{a}	.45**	.29	-.03	.37*	-.21

Notes- ^{a} *one-tailed p < .05 * two-tailed p < .05 ** two-tailed p < .01*

Correlations were computed listwise (df = 35)

Table 17. Correlations between predictors and number of revisits on novel label trials in Experiment 3.

	CHOICES					
PREDICTOR	3	4	5	6	Overall Mean	Linear Trend
1. Age	.18	.07	.32[a]	.30[a]	.34*	.06
2. Word Knowledge	-.02	-.02	.32[a]	.09	.12	-.09
3. Object Nameability	.15	-.04	.31[a]	.01	.21	-.05
4. Vocabulary Size	.04	-.08	.21	.28[a]	.16	.07

*Notes- [a] one-tailed p < .05 * two-tailed p < .05 ** two-tailed p < .01*

Correlations were computed listwise (df = 35)

Table 18. Correlations between predictors and proportion of looking time spent on the target on novel label trials in Experiment 3.

	CHOICES					
PREDICTOR	3	4	5	6	Overall Mean	Linear Trend
1. Age	.04	-.08	-.36*	-.02	-.15	-.09
2. Word Knowledge	-.10	.10	-.31[a]	-.01	-.11	-.04
3. Object Nameability	.04	.07	-.19	-.01	-.03	-.08
4. Vocabulary Size	-.11	.04	-.30[a]	-.26	-.24	-.20

Notes- [a] *one-tailed* $p < .05$ * *two-tailed* $p < .05$ ** *two-tailed* $p < .01$

Correlations were computed listwise (df = 35)

Table 19. Correlations between predictors and number of foils checked on familiar label trials in Experiment 3.

	CHOICES					
PREDICTOR	3	4	5	6	Overall Mean	Linear Trend
1. Age	.26	.06	.13	.20	.22	.06
2. Word Knowledge	.00	.14	.26	.02	.14	.03
3. Object Nameability	.06	-.04	.15	-.11	.00	-.10
4. Vocabulary Size	.23	.17	.17	-.04	.15	-.18

Notes- ^{a} *one-tailed* $p < .05$ ** two-tailed* $p < .05$ *** two-tailed* $p < .01$

Correlations were computed listwise (df = 35)

Table 20. Correlations between predictors and number of revisits on familiar label trials in Experiment 3.

PREDICTOR	CHOICES 3	 4	 5	 6	Overall Mean	Linear Trend
1. Age	.01	-.02	.33*	.00	.12	.06
2. Word Knowledge	.12	-.13	.23	-.14	.01	-.09
3. Object Nameability	-.07	-.16	.08	-.22	-.17	-.05
4. Vocabulary Size	.11	-.16	.35*	.06	.13	.07

Notes- ^{a} *one-tailed* $p < .05$ * *two-tailed* $p < .05$ ** *two-tailed* $p < .01$

Correlations were computed listwise (df = 35)

Table 21. Correlations between predictors and proportion of looking time spent on the target on familiar label trials in Experiment 3.

arrow	blender	chair	fan	headphones	net	scissors	thermometer	whistle
backpack	blocks	computer	felt-tip marker	house	pen	shoe	tissues	window
ball	boat	cookie	fire truck	house key	pencil	shopping cart	toaster	zipper
balloon	book	couch	flashlight	jacket	piano	shorts	toilet	
band-aid	bottle	crayon	flip flops	knife	pillow	sink	toothbrush	
baseball bat	bowl	crown	football	lamp	plane	skateboard	toothpaste	
basket	bottom	cup	fork	lightbulb	plate	socks	towel	
bathtub	bucket	diaper	garbage can	mailbox	potato chips	spoon	trampoline	
bed	button	door	gloves	microphone	present	stapler	umbrella	
bell	calendar	doorknob	guitar	microwave	refrigerator	t-shirt	vacuum	
belt	camera	dress	hairbrush	mop	rubber band	table	violin	
bib	candle	eraser	hammer	necklace	scarf	target	wagon	
bicycle	car	eyeglasses	hat	necktie	school bus	television	watch	

Table 22. Names of objects which pictures were used as familiar stimuli in Experiment 1.