Running Head: Phonological similarity

Linguistic Labels: Conceptual Markers or Object Features?

Vladimir M. Sloutsky

Department of Psychology and Center for Cognitive Science

The Ohio State University

Anna V. Fisher

Department of Psychology

Carnegie Mellon University

July 20, 2011

In press: *Journal of Experimental Child Psychology*

Please address correspondence to:

Vladimir M. Sloutsky
Center for Cognitive Science
208C Ohio Stadium East
1961 Tuttle Park Place
The Ohio State University
Columbus, OH 43210
Phone: (614) 688-5855
Fax: (614) 292-0321
Email: Sloutsky.1@osu.edu

ABSTRACT

Linguistic labels affect inductive generalization; however, the mechanism underlying these effects remains unclear. According to one similarity-based model (SINC**:** **S**imilarity-**I**nduction-**N**aming-**C**ategorization), early in development labels are features of objects contributing to the overall similarity of compared entities, with early induction being similarity-based. If this is the case, then not only identical, but also phonologically similar labels may contribute to the overall similarity, and therefore to induction. These predictions were tested in a series of experiments with 5-year-olds and adults. In Experiments 1-5, participants performed a label extension task, whereas in Experiment 6 they performed a feature induction task. Results indicate that phonological similarity contributes to early induction, and support the notion that for young children labels are features of objects.

Linguistic Labels: Conceptual Markers or Object Features?

Words play an important role in directing inductive generalization from known to novel. For example, when two appreciably different entities are referred to as “dogs” and one of the dogs is described as having a particular property (e.g., it has short bones), even young children are more likely to generalize this property to another dog than when the entities are referred to as “a dog” and “a cat”, or when no words are introduced (Gelman & Markman, 1986; Sloutsky, Lo, & Fisher, 2001; Sloutsky & Fisher, 2004). However the mechanism by which words guide young children’s induction remains unclear.

Some researchers suggest that young children hold a number of *conceptual* assumptions about the world and language (see, Gelman & Coley, 1991; Keil, Smith, Simons, & Levin, 1998; Murphy, 2002, for reviews of these assumptions), and words guide induction by invoking this conceptual knowledge. First, young children assume that (a) individuals belong to categories and (b) things belonging to the same category have much in common. Second, they assume that linguistic labels presented as count nouns are symbols denoting categories. Thus, on the basis of these assumptions, young children conclude that things having the same name are likely to have much in common.

Another theoretical position argues that there is no need to posit rich conceptual knowledge to understand effects of words on young children’s induction. For example, a similarity-based model of generalization in children (SINC -- for Similarity, Induction, Naming, and Categorization -- Sloutsky & Fisher, 2004) argues that (a) young children perform induction on the basis of the overall similarity among compared entities, (b) shared linguistic labels are features contributing to the overall similarity, and (c) the process of computing similarity over visual and auditory features is automatic rather than deliberate (see Sloutsky & Fisher, 2005, for discussion). In the next section, we present SINC in greater detail.

SINC: Labels as Features Contributing to Similarity

SINC considers labels as features contributing to the overall similarity of compared entities (Sloutsky & Fisher, 2004; Sloutsky & Lo, 1999, Sloutsky, et al. 2001, see also Sloutsky & Fisher, 2005, for a discussion). There is evidence supporting this assumption of SINC: when entities share a label, young children tend to consider these entities as looking more alike than when the same entities are presented without labels (Sloutsky & Fisher, 2004; Sloutsky & Lo, 1999). According to the theory underlying SINC, these effects stem from attentional factors, such as auditory information overshadowing (or attenuating processing of) corresponding visual information (Napolitano & Sloutsky, 2004; Robinson & Sloutsky, 2004; 2007a, 2007b, 2008**;** Sloutsky & Napolitano, 2003).

Initial evidence for overshadowing was presented in the Sloutsky and Napolitano (2003) study, in which 4-year-olds and adults were presented with an auditory-visual target item followed by a test item. Participants had to respond *same* if the two compound stimuli had the same auditory and visual components, and to respond *different* if either the auditory or visual component differed between the target and test items. The auditory components consisted of unfamiliar nonlinguistic sounds and the visual components consisted of unfamiliar images (e.g., abstract geometric shapes). If participants encode both auditory and visual stimuli they should correctly accept target items as the same, while correctly rejecting items that had either new visual or new auditory components as different. It was found that 4-year-olds failed to report that the visual components changed when visual input was accompanied by auditory input. At the same time, processing of visual stimuli was not difficult per se: in the absence of auditory input, young children ably encoded the visual input (Napolitano & Sloutsky, 2004; Robinson & Sloutsky, 2004; Sloutsky & Napolitano, 2003). Similar effects were found in 8-, 12-, and 16-month-old infants (Robinson & Sloutsky, 2004).

To introduce SINC more formally, we will first consider its predictions for a class of frequently used induction tasks (Gelman & Markman, 1986; Sloutsky & Fisher, 2004). In these tasks, the child is presented with a Target item and two Test items, A and B. After labels for the Target and Test items are introduced, the child is taught that Test A has a particular quasi-biological property (e.g., has hollow bones) and Test B has a different quasi-biological property (e.g., has solid bones). The task is to decide whether the Target has hollow bones like Test A or solid bones like Test B. SINC predicts that the probability of inducing a property from a test item (say Test B) to the Target is a function of ratio of the overall similarity of Test A to the Target and Test B to the Target, which is a consequence of Luce's choice rule. More formally, this probability is presented in Equation 1.

$$P(B) = \frac{Sim(B,T)}{Sim(B,T) + Sim(A,T)} = \frac{1}{1 + \frac{Sim(A,T)}{Sim(B,T)}} \qquad (1)$$

where *Sim* (A, T) and *Sim* (B, T) are similarities of Test A and Test B to the Target, respectively. Furthermore, according to SINC, the overall similarity between each of the Test items and the Target is a function of visual similarity and the weight of linguistic label.

$$Sim(B,T) = \lambda \upsilon^{B} \qquad (2)$$

where λ and υ (these parameters vary between 0 and 1) denote values (weights) of a label and visual properties, respectively, while *B* denotes the number of featural mismatches between the appearance of Test B and the Target. When there are no mismatches, *B* = 0, and λ = υ =1, thus the similarity between Test B and the Target equals to unity.

Now substituting similarities in Equation 1 with those in Equation 2, we can express the probability of inducing a property from Test B to the Target.

$$P(B) = \frac{1}{1 + \lambda \frac{\upsilon^{A}}{\upsilon^{B}}} \qquad (3)$$

where λ is weight of label mismatch, and $\frac{\upsilon^A}{\upsilon^B}$ is the ratio of appearance similarities of each of the test items to the Target. Therefore, according to SINC, the probability of inducing a property from Test B to the Target is a function of the ratio of appearance similarities of each of the Test items' to the Target's appearance, and the weight of label.

There are a number of reports supporting predictions of SINC. In particular, Sloutsky & Fisher (2004) demonstrated that SINC accurately predicts induction and categorization performance of young children with a variety of stimuli, including those that were previously used in research supporting the naïve theory approach to induction (such as picture triads used by Gelman and Markman, 1986). In addition, as mentioned above, there is evidence supporting the basic assumption of SINC that early in development labels are features contributing to the overall similarity of compared entities. In particular, when two entities share a label, young children tend to consider these entities as looking more alike than when the same entities are presented without labels (Sloutsky & Fisher, 2004; Sloutsky & Lo, 1999).

However, previous research testing SINC (e.g., Sloutsky & Lo, 1999; Sloutsky et al., 2001; Sloutsky & Fisher, 2004) used only triads in which items had either identical or different labels, with identical labels contributing to similarity and thus to induction. At the same time, if labels are features contributing to similarity, it is possible that labels are perceived as subjectively continuous variables, in which case, not only the identity, but also phonological similarity of labels would contribute to the overall similarity and to induction (Equation 4).

$$Sim(B,T) = \lambda^{\beta}\upsilon^{B} \tag{4}$$

where λ and υ denote values (weights) of a label and visual property, respectively, while B and β denote the number of featural mismatches between the appearance and the label of Test B and the

Target, respectively. When there are no mismatches, $\beta = B = 0$, and $\lambda = \upsilon = 1$, the similarity between Test B and the Target equals to unity. Finally, substituting similarities in Equation 4 with those in Equation 1, we can express the probability of inducing a property from Test B to the Target.

$$P(B) = \frac{1}{1 + \frac{\lambda^{\alpha}}{\lambda^{\beta}} \frac{\upsilon^{A}}{\upsilon^{B}}}$$

where λ^{α} is the similarity of label of Test A to the label of the Target, whereas λ^{β} is the similarity of label of Test B to the label of the Target. Therefore, the probability of inducing a property from Test B to the Target is a function of two similarity ratios: (a) the ratio of appearance similarities of each of the Test items to the Target's appearance and (b) the ratio of label similarity of each of the Test items to the Target's label.

There is evidence supporting (albeit indirectly) the possibility of effects of phonological similarity on induction. In particular, Merriman and Schuster (1991) demonstrated that phonological similarity of words led to attenuated "mutual exclusivity" effects in 2- and 4-year-olds. Mutual exclusivity is a tendency of toddlers and young children to interpret novel words as referring to novel rather than to familiar items (Markman & Wachtel, 1988). For example, when presented with an apple and a novel object and asked to select a "Dax," participants pointed to the novel object. Merriman and Schuster found that the effect reduces markedly or disappears if the novel word is phonologically similar to a familiar word: children would not necessarily extend a novel word "Japple" to a novel object, but may rather extend it to an apple.

Although these findings might be indicative of the fact that phonological similarity of words is important for young children, there is another possible interpretation of these findings (and Merriman and Schuster explicitly pointed to this possibility). In particular, these findings may stem from young children's tendency to interpret phonologically similar words as variants of the target words. First, it is possible that young children interpret a phonologically similar variant of a familiar target word as a morphological derivation of the target word (i.e., *Japple* is to *Apple* is like *Doggie* is to *Dog*). Another possibility is that children interpret a phonologically similar word as an alternative pronunciation of the target word (similar to alternative pronunciations of words such as *route* and *data*) or as a different speech token of the target word.

However, there is evidence challenging these alternative interpretations of Merriman and Schuster's (1991) findings, while supporting the possibility that phonological similarity is important for early generalization. According to a report by Dewhurst and Robinson (2004), even when phonologically similar words are known to be different words (e.g., *Bag* vs. *Bad*), phonological similarity affects the way words are represented in memory. These authors demonstrated that young children are more susceptible to phonological errors on the recall tasks than older participants. In particular, young children falsely reported that the word *Bag* was presented, whereas in fact the presented word was *Bad*. At the same time, preadolescents and adults are more prone to semantic intrusions (i.e., they falsely report that the word *Good* was presented, whereas in fact the presented word was *Bad*). These effects could not stem from young children interpreting phonologically similar words as morphological variants or alternative pronunciations of the same word because all words were familiar to young children.

The Merriman and Schuster (1991) study in conjunction with the Dewhurst and Robinson (2004) study present indirect evidence supporting the prediction of SINC that phonological similarity of labels may affect induction in young children. The goal of the current research is to directly test this prediction. The second goal was to test another prediction that follows directly from the idea that early in development phonological similarity of labels contributes to the overall similarity and thus to induction. This prediction concerns performance on a lexical extension task (i.e., “this is a *gatu*, show me a *zatu*”). Since label extension is a variant of an induction task, with participants generalizing a label rather than a property, performance on a label extension task should also be driven by the overall similarity, such that children should map phonologically similar labels to visually similar items. There is some evidence supporting this prediction: young children extend identical words (with identity being the maximal similarity) to similarly looking items (Gentner, 1978), however effects of phonological similarity on label extension in young children have never been examined directly.

Both predictions (i.e., that phonological similarity contributes to the induction of properties and to label extension) are novel, and they follow from the central assumption of the model. The qualitative predictions of SINC pertaining to label extension (i.e., that phonologically similar labels get mapped onto similarly looking items) were tested in Experiments 1-5, whereas quantitative predictions of SINC pertaining to induction of properties (i.e., to what extent phonological similarity of labels contributes to induction) were tested in Experiment 6.

EXPERIMENT 1

The goal of Experiment 1 was to test the prediction of SINC that phonological similarity of labels directs young children’s induction. Experiment 1 used a variant of an induction task, in which young children extended the label of the target item to one of the test items. If phonological similarity guides early induction, then young children should extend phonologically similar labels to perceptually similar entities. At the same time, this may not be the case for adults, who may treat labels as proxies for category membership rather than object features (Sloutsky & Lo, 1999; Sloutsky & Fisher, 2004; Yamauchi & Markman, 2000).

Method

Participants

Participants were 18 preschool-age children (10 girls and 8 boys; *M* = 5.26 years; *SD* = 0.33 years), recruited from childcare centers located in middle class neighborhoods of the Columbus Ohio area. In this and all other experiments reported here, participants were included in the sample on the basis of returned parental consent forms. There were also 15 adults (5 women and 10 men, *M* = 21.21, *SD* = 3.45) participating in this experiment in exchange for a partial course credit.

Materials

Materials consisted of eight sets of color photographs and eight pairs of artificial labels calibrated for phonological similarity. Calibration was done in a separate experiment with 103 young children (see appendix for details of the calibration experiment), in which participants had to decide which of the two test words sounded more similar to the target. The calibration experiment yielded three levels of phonological similarity, with some test items being identical to the target (e.g., both were referred to as *gama*), some being similar to the target (e.g., one test item being referred to as *gatu* and the target being referred to as *zatu*), and some being different from the target (e.g., one test item being referred to as *satu* and the target being referred to as *kipa*). Special care was taken to ensure that phonological similarity would not communicate morphological relatedness (see appendix for details). Experiments 1-5 used either phonologically similar or phonologically different labels, whereas Experiment 6 used labels at all three levels of phonological similarity. The list of label pairs used in Experiments 1-5 is presented Table 1A.

While being phonologically similar, these labels were discriminable by young children. The discriminability of labels was established in a calibration experiment (see Experiment 5 for a detailed description) with a separate group of 20 children. These participants were asked to judge whether two successively presented labels were identical or not, with half of these label pairs being identical and half being phonologically similar. The results indicated that young children had little difficulty discriminating phonologically similar labels (d-prime = 3.41, above 0, one sample t (19) = 9.76, $p < .0001$). Therefore, (if found) phonological similarity effects cannot be attributed to children's inability to discriminate phonologically similar labels.

Each picture set consisted of colored photographs of a Target (e.g., red grapes) and four Test items: Test 1 was identical to the Target (i.e., Identical), Test 2 depicted a similar entity from the same basic level category as the Target, such as green grapes (i.e., Similar), Test 3 depicted an entity coming from the same superordinate category as the Target, such as cherries (i.e., Less Similar), and Test 4 was perceptually and ontologically different from the Target, such as a purse (i.e., Dissimilar). Examples of pictures sets are presented in Figure 1.

Procedure

In this and all other experiments presented below, children were tested individually in their childcare centers and adults were tested individually in a laboratory on campus by hypothesis-blind female experimenters. Picture sets were presented on a computer screen. Participants were told that they would see pictures of things that they may know and learn what these things are called "in a far away place". There were a total of 8 trials. On each trial a Target was labeled with an artificial word, and participants were asked to find a Test item that would be labeled with a phonologically similar word. For instance, participants could be told that the Target is called a *gama* in a far away place and asked to find a Test item that is called a *guma* in the far away place. Position of Test items on the screen and the order of trials were randomized.

Results and Discussion

Proportions of label extensions across the four levels of Test-Target similarity for children and adults are presented in Figure 2 (these proportions for children across experiments are also presented in Table 1B). Data in the Figure indicate that young children overwhelmingly extended phonologically similar labels to identical items (i.e., 84%, above chance, one-sample t (17) = 8.19, Bonferroni-adjusted $p < .0001$), while rarely selecting the Similar, Less Similar, and Dissimilar items (i.e., 9%, 4%, and 3% for respectively). Furthermore, out of 18 tested children, 16 demonstrated a dominant pattern of responding (i.e., selecting the same type of test item on at least 75% of the trials, or at least 6 out of 8 trials). In contrast, adults' responses were more diffused, with only 4 out of 15 participants demonstrating a dominant pattern of responding (two participants chose Similar test items, one participant chose Less Similar items, and one participant chose Dissimilar items consistently) and none of the choice options eliciting above chance responding (all Bonferroni-adjusted ps > .12). The difference in the proportion of children and adult participants demonstrating a dominant pattern of responding was significant, $\chi^2(1, N = 33) = 13.27$, $p < .001$.

These results point to an important difference between label extension in young children and adults. Whereas young children overwhelmingly extend similar labels to identical items, adults exhibit no dominant pattern of responding, with their responses being diffused across the choice options. These results indicate that phonological similarity of labels contributes to young children's generalization (something that could be expected if labels are treated as continuous features), while it has no contribution to generalization of adults (something that could be expected if labels are treated as category markers).

It could be argued, however, that young children were merely matching pictures without attending to the phonological similarity of labels. According to this explanation, even when presented with different labels, children should continue to match pictures and extend different words to visually similar entities. This possibility was tested in Experiment 2, in which children were presented with the same visual stimuli as in Experiment 1, except that this time phonologically different labels were used (see Table 1A for examples).

Experiment 2

Method

Participants

Participants were 14 preschool-age children (11 girls and 3 boys; $M = 4.92$ years; $SD = 0.26$ years), recruited from childcare centers located in middle class neighborhoods of the Columbus Ohio area.

Materials and Procedure

Materials and procedure were similar to those of Experiment 1, with one important difference: phonologically different labels, such as *kipa*-*satu,* were used instead of similar labels (see Table 1A for the list of labels used in Experiment 2).

Results and Discussion

Results of Experiment 2 differed markedly from those of Experiment 1 (see Table 1B). In particular, when presented with phonologically different labels young children were equally likely to extend these labels to each of the four types of test items (28%, 26%, 25%, and 21% of label extensions to Identical, Similar, Less Similar, and Dissimilar test items respectively, not different from chance, all *p*s > .45). Taken together with results of Experiment 1, current findings indicate that children exhibit different patterns of responses when extending phonologically similar labels and phonologically different labels. Therefore, it seems highly unlikely that children's responding in Experiment 1 was driven by their tendency to merely match pictures.

Note that in Experiment 1 children overwhelmingly extended similar labels to Identical items with identity representing the maximum level of similarity. Therefore if children's induction is guided by similarity and is not limited to identity, then removing identical items from the test set should encourage young children to extend phonologically similar labels to the next most similar to the Target item in the set (i.e., the Similar item). To test this prediction, we conducted Experiment 3, in which we presented participants with the task and materials identical to the ones used in Experiment 1 with one important difference: items identical to the Target were removed from the sets of test items.

EXPERIMENT 3

Method

Participants, *Materials*, *and Design*

Participants were 15 preschool-age children (10 girls and 5 boys; $M = 4.99$ years; $SD = 0.35$ years), recruited from childcare centers located in middle class neighborhoods of the Columbus Ohio area. Materials and procedure were similar to those of Experiment 1, however the Identical items were removed from the test sets.

Results and Discussion

Results of this experiment are presented in Table 1B. When Identical items were removed from the test set, young children overwhelmingly extended phonologically similar labels to Similar items (i.e., 78%, above chance, one-sample t (14) = 10.52, Bonferroni-adjusted $p < .0001$), whereas they rarely extended similar labels to either Less Similar or Dissimilar test items (i.e., 19% and 3%, respectively). These findings replicated and further extend results of Experiment 1: young children extend phonologically similar labels not only to identical items (as in Experiment 1) but also to visually similar items.

However, it could be argued that children generalized phonologically similar labels along the lines of visual similarity in Experiments 1 and 3 because they considered similar labels to be variants of pronunciation of the same words, similar to different variants of pronunciation of such words as *route* and *data*. We consider this possibility to be highly unlikely for three reasons. First, words with alternative pronunciations are highly infrequent. An extensive search that included two linguists generated just five such words in American English: route, data, tomato, aunt, and vase. Given such infrequency, the hypothesis that words can have alternative pronunciation should be highly counterintuitive for a young child. In addition, there is no evidence indicating that young children know both pronunciations for these words. And finally, results of Experiment 1 provide little evidence that even adults generate such a counterintuitive hypothesis for phonologically similar words.

Another, more likely, possibility is that children may consider the two phonologically similar words as different phonological tokens of the same word. This possibility is more likely because every word is pronounced differently by different speakers, yet even young infants have little difficulty ignoring this surface variation and perceive the variants as phonological tokens of the same word (see Jusczyk, 1997 for a review, see also Swingley & Aslin, 2007).

We addressed this possibility in Experiment 4 in which we presented participants with the same label extension task used in Experiments 1 and 3; however instead of using familiar objects and novel label pairs, in Experiment 4 we presented participants with novel objects and familiar phonologically similar labels (i.e., *mouse* - *house*). There is little doubt that children do not treat phonologically similar familiar labels, such as “house” and “mouse”, as different tokens of the same word. Therefore, if we replicate the findings of Experiment 1 with phonologically similar familiar labels, it would render the possibility that these results stemmed from children interpreting phonologically similar words as tokens of the same word highly unlikely.

EXPERIMENT 4

Method

Participants

Participants were 22 preschool-age children (11 girls and 11 boys; $M = 4.9$ years; $SD = 0.42$ years), recruited from a childcare center located in a middle class neighborhood in Pittsburgh, Pennsylvania.

Materials and Design

Materials consisted of eight sets of color pictures of unfamiliar objects, half of which represented novel artifacts and another half represented novel natural kind-like items (see Figure 1 for examples), and eight label pairs consisting of familiar phonologically similar labels (see Table 1A for the list of label pairs used). A subset of children ($N = 11$) who participated in Experiment 4 was interviewed one week after the completion of the experiment proper to establish that labels used in Experiment 4 were indeed familiar to 5-year-old children. Participants were presented with 16 displays consisting of four pictures each, and asked to identify objects corresponding to the labels used in Experiment 4. Children were over 99% accurate in performing this task.

Results and Discussion

Results of this experiment are presented in Table 1B. The pattern of label extension obtained with familiar similar labels (which children are unlikely to interpret as variants of pronunciation of the same words) in Experiment 4 was similar to the pattern of results obtained with similar novel labels in Experiment 1. In particular, young children extended phonologically similar labels to identical items (i.e., 65%, above chance, one-sample t (21) > 4.57, Bonferroni-adjusted $p < .001$, $d = 1.33$), whereas they rarely extended similar labels to Similar, Less Similar, and Dissimilar objects (i.e., 10%, 12%, and 13% respectively). The use of familiar labels undermines the possibility that young children interpreted phonologically similar words (e.g., *mouse* and *house*) as tokens of the same word. Given that familiar phonologically similar words used in the current experiment elicited a similar pattern of responses as novel phonologically similar (yet discriminable) words used in Experiments 1 and 3, it seems likely that label extension in Experiments 1 and 3 was driven by phonological similarity of labels rather than young children considering phonologically similar labels to be alternative pronunciations of the same word.

Overall, results of Experiments 1-4 (these results are presented in Table 1B) generate several important findings. First, label extension in young children is driven by similarity: young children extend phonologically similar labels (but not dissimilar labels) to visually similar items. Moreover, effects of phonological similarity on label extension are strong even when children are presented with novel objects and familiar similar labels. Given that the words used in this experiment were highly familiar (e.g., “mouse”-“house” or “flower”-“shower”), it is very unlikely that children considered the words pairs as phonological tokens of the same word. Therefore, results of Experiment 4 strongly support the phonological similarity hypothesis, while undermining alternative interpretations of results of Experiments 1 and 3.

In sum, results of Experiments 1-4 suggest that for young children phonological similarity of labels affects label extension, with similar sounding labels being extended to similar looking entities. If this is the case, then reducing phonological similarity of labels should also affect the pattern of label extension. The goal of Experiment 5 was to test this hypothesis, thus providing converging evidence for findings of Experiments 1-4.

EXPERIMENT 5

In Experiment 5 young children were presented with a label extension task identical to that in Experiment 1, however the experiment proper was preceded by discrimination training. Note that labels pairs used in Experiments 1-3 were discriminable, yet phonologically similar. At the same time, it was assumed that perceptual similarity is a function of discriminability, such that the higher the discriminability, the lower the perceived similarity. Therefore, a training procedure that increases discriminability should decrease similarity of phonologically similar labels. The goal of Experiment 5 was to provide such discrimination training and to examine its effects on phonological similarity and on the pattern of label extension.

Discrimination training was followed by the label extension task used in Experiment 1. There were two training conditions: Experimental and Control. In the Experimental condition, the trained label pairs were also used in the label extension task. In the Control condition, participants were trained on one set of phonologically similar pairs, whereas a different set of phonologically similar pairs was used in the label extension task.

The experiment was based on the following logic. If young children's performance in Experiments 1 and 3 was driven by phonological similarity, then discrimination training (which should result in decreased phonological similarity of presented labels) should decrease the level of label extensions to identical items in the Experimental, but not in the Control condition.

Note that training was focused strictly on discriminability (and thus on similarity) of labels, and care was taken to ensure that young children would not consider training as a hint as to whether the phonologically similar words are token pronunciations of the same word or two different words. In particular, the discrimination task differed from the experiment proper and the words "same," "different," and "word" were never mentioned in the course of training. Furthermore, in terms of potential hints, the Experimental and Control conditions were identical. Therefore, any considerations prompted in young children by the Experimental condition should be also prompted by the Control condition.

Method

Participants

Participants were 38 preschool-age children (21 girls and 17 boys; *M* = 5.14 years; *SD* = 0.49 years), with 17 children participating in the Experimental training condition and 21 children participating in the Control training condition.

Design, Materials, and Procedure

Experiment 5 consisted of two parts: discrimination training and the experiment proper. There were two discrimination training conditions, Experimental and Control, with both conditions having the same procedure, but different training materials. In the Experimental condition, training and the experiment proper used similar materials: labels used in the Experimental condition were a subset of labels subsequently presented to participants in the experiment proper (these were labels used in Experiment 1). In particular, the following six label pairs were used in the course of training in the Experimental condition[1]: (1) *zatu – gatu*, (2) *pika – fika*, (3) *sudu – su-gu,* (4) *gama – guma,* (5) *tabi – da-bi,* (6) *fima – fina.* The Control condition was similar to the Experimental condition in length and task demands, however, the label pairs used in the Control condition were unrelated to the label pairs used in the experiment proper (e.g., *tibi – tobi, tudi – tugi*, *zuma – zu-uma*, etc.).

Discrimination Training. During discrimination training, participants were told that they would play a guessing game, in which they would see animal cartoons presented on a computer screen. The cartoons could appear either on the left or on the right side of the computer screen. The experimenter then demonstrated how children can predict which cartoon will appear and where. For example, participants were told that when they hear the computer play [at this time *zatu* was played by the computer], a dog cartoon would appear on the left side of the screen, whereas when they hear the computer play [at this time *gatu* was played by the computer], a bird cartoon would play on the right side of the screen. Participants were then told that their task was to predict on which side of the screen the cartoon would appear by carefully listening to the computer. During discrimination training, participants received feedback: cartoons appeared only following correct predictions, whereas incorrect predictions were followed by an unrelated still picture appearing at the center of the screen. The discrimination task consisted of six training blocks (one block per label pair), with 10 trials per block, and with the total of 60 trials. The order of blocks was constant for all participants, and the order of trials within a block was randomized for each participant. At the conclusion of training in both Experimental and Control conditions participants were praised for their performance and presented with the experiment proper, which was identical to Experiment 1.

Effect of Discrimination Training on Phonological Similarity. The ability of training to attenuate phonological similarity by increasing discrimination was tested in a separate calibration experiment with 42 4-5 year-olds none of whom participated in previous experiments or in Experiment 5. There were two between-subjects conditions, with 20 children participating in a baseline discrimination condition and 22 participating in the training condition. In the baseline condition, participants were presented with a discrimination task, whereas in the training condition, the discrimination task was preceded by the described above training procedure. The discrimination task included 8 label pairs used in Experiment 2 and 8 pairs of identical labels. Participants' goal was to determine whether the pair consisted of the same words or different words. Discrimination accuracy was calculated as difference between the proportion of hits (correct detection that two presented labels were different) and false alarms (incorrect reporting that two identical labels were different). When there is no discrimination, hits equal false alarms and discrimination accuracy equals to 0. Perfect discrimination results in accuracy of 1. Results of the calibration experiment indicated that participants reliably discriminated phonologically similar labels in the baseline condition (Hits = .71, False Alarms = .09, accuracy = .62, above 0, one sample t (19) = 12.37, $p <$ .0001). More importantly, training resulted in significantly improved discrimination of the phonologically similar labels, (Hits = .82, False Alarms = .06, accuracy = .76, above the baseline, independent-sample, t (40) > 2.09, $p < .05$). Given that training increases discrimination of phonologically similar labels, it was expected that in Experiment 5 training would somewhat attenuate phonological similarity of labels (by increasing their discriminability). The attenuation of phonological similarity of labels would in turn affect the pattern of label extension in the Experimental condition, which used the same label pairs for training and testing, but not in the Control condition, which used different label pairs for training and testing.

Results and Discussion

During discrimination training, nine participants (four in the Experimental and five in the Control condition) responded accurately on less than 75% of trials and their data were excluded from further analysis. The rest of participants averaged 86% of correct responses in the Experimental condition, and 87% of correct responses in the Control condition, above chance, both one-sample ts > 17.69, $p < .0001$, and our analysis of the label extension data is limited to these participants.

Proportions of label extensions in the Experimental and Control conditions, as well as proportions of label extensions without training (Experiment 1) are presented in Figure 3. As can be seen in the figure, discrimination training in the Experimental condition changed the patterns of label extension. To analyze this change, proportions of label extensions to Identical items in Experiments 1 and 5 were submitted to a one-way ANOVA with the training condition as a factor (no training in Experiment 1, Experimental training condition in Experiment 5, and Control training condition in Experiment 5). Results of the analysis revealed significant differences among these conditions (F (2, 44) = 4.29, p < .05, η_p^2 = .16), and a post-hoc Tukey test revealed that the decrease in proportion of label extensions to identical items in the Experimental condition of Experiment 5 (50%) was significant compared to Experiment 1 (84%), Cohen's d = .87, and the Control condition of Experiment 5 (83%), Cohen's d = .84, both ps < .05. At the same time, the proportions of extensions in the Control condition of Experiment 5 (83%) and Experiment 1 (84%) were statistically equivalent (p > .99). These findings clearly indicate that effects observed in the Experimental condition of Experiment 5 stemmed from discrimination training resulting in attenuated similarity of labels, and not from extraneous factors.

The results of Experiment 5 demonstrate that increasing discriminability (and thus decreasing phonological similarity) of labels has a strong effect on children's performance on a label extension task. In particular, children were less likely to extend phonologically similar labels to identical items after being trained to discriminate pairs of labels used in a subsequent label extension task than after being trained to discriminate pairs of labels irrelevant to the subsequent task. These results provide converging evidence that phonological similarity of labels affects label extension in young children.

Taken together, results of Experiments 1-5 indicate that phonological similarity of labels affects generalization in young children. Furthermore, Experiments 4 and 5 undermine alternative interpretations of the phonological similarity effect, and we come back to this issue in the General Discussion section. Whereas results of Experiments 1-5 support predictions of SINC pertaining to the label extension task, and the goal of Experiment 6 was to test predictions of SINC pertaining to the property induction task.

EXPERIMENT 6

To examine effects of phonological similarity on induction of properties and to test predictions of SINC, Experiment 6 used a variant of the triad task, which includes a Target and two Test items, with one Test item being more similar to the Target and the other having a more similar label. Participants were told that Test A had a particular property (e.g., hollow bones), whereas Test B had another property (e.g., solid bones) and their task was to decide whether the Target had the same property as Test A or as Test B. To separate the contribution of phonological similarity from the contribution of visual similarity, we independently varied both the ratio of visual similarity of entities and the ratio of phonological similarity of labels.

Although the ratio of similarities is a continuous variable, an experiment can only have a finite number of trials, and, particularly when participants are young children, this number must be rather small. Therefore, similarity ratios were broken down into two levels for the visual dimension (see Figure 4), and three levels for the auditory dimension (see Figure 5). As shown in Figure 4, at Level 1, appearances were not predictive, with the Target and both Test items being identical, and similarity ratio was equal to 1. At Level 2, appearances were fully predictive, with Test A being identical to the Target, and Test B being very dissimilar. In this case, the similarity ratio (i.e., Sim (A, Target)/Sim (B, Target)) was equal to 9, with Test A, but not Test B being similar to the Target.

There were three levels of phonological similarity (see Figure 5). At Level 1, labels were not predictive, with Test and Target labels being the same, and similarity ratio was equal to 1. At Level 2, labels were partially predictive, with the label of Test B being somewhat more similar to the Target's label than the label of Test A. At Level 2, the label of Test B was more similar to the Target label than that of Test A, and therefore the ratio was somewhat smaller than 1 (i.e., Sim (A, Target)/Sim (B, Target) was equal to 0.42). Finally, at Level 3, labels were fully predictive, with

Test B label being identical to the Target label, and Test A label being very dissimilar. In this case, the ratio was much smaller than 1 (i.e., Sim (A, Target)/Sim (B, Target) was equal to 0.18).

To discern the contribution of labels and appearances, Test A always had the same or greater appearance similarity to the Target than Test B (thus the ratio of appearance similarities was equal to or greater than 1), whereas it had the same or smaller label similarity to the Target than Test B (thus the ratio of label similarities was equal to or smaller than 1). Proportion of Test B choices predicted by SINC (derived by plugging the respective similarity ratios into Equation 5) are presented in Figure 6A. As shown in the figure, SINC predicts the main effect of visual similarity on induction: the probability of inducing a property from Test B to the Target would decrease with a decrease in visual similarity of Test B to the Target. This prediction was tested and confirmed previously (e.g., Sloutsky & Fisher, 2004). More importantly, SINC predicts the main effect of phonological similarity on induction: the probability of inducing a property from Test B to the Target should increase with an increase of phonological similarity of Test B's label to the label of the Target. In other words, the more similar the labels of Test B and the Target (and thus the smaller the similarity ratio of labels) are, the more likely the induction of properties from Test B to the Target. The goal of Experiment 6 was to test this prediction of SINC.

Method

Participants

Participants were 67 preschool children (31 girls and 36 boys; M = 4.9 years; SD = 0.34), with 20 participants in the Phonological Ratio = 1, 22 participants in the Phonological Ratio = 0.42, and 25 participants in the Phonological Ratio = 0.18 conditions.

Materials and Design

Materials consisted of visual and auditory stimuli, which were calibrated for similarity. Visual stimuli were 8 triads of pictures of animals selected from the sequences of images, in which one animal was "morphed" into another in a fixed number of steps, in the manner described in Sloutsky and Fisher (2004). Visual similarity ratios were estimated in a separate calibration experiment (see Sloutsky & Fisher, 2004, for the details of calibration). During calibration, young children were presented with calibration triads consisting of a Target and two test items, Test A and Test B, such that Test A was either equally or more similar to the Target compared to Test B. Participants were then asked whether the Target looked more like Test A or Test B, and proportions of Test A and Test B selections were converted into similarity ratios (similarity ratio = proportion of Test A selections / proportion of Test B selections). When Test A and Test B are non-discriminable, visual similarity ratio equals 1, and as discriminability of test items increases, so does the visual similarity ratio. There were two levels of visual similarity with 4 triads per level: Ratio = 1 (both Test stimuli looked equally similar to the Target) and Ratio = 9 (Test A looked almost identical to the Target, while Test B looked very different).

Auditory stimuli were selected from 36 pronounceable two-syllable non-words recorded by a male native English speaker, each ranging from 560 to 1120 ms in duration. These stimuli were calibrated in a separate experiment with 103 young children (see Appendix for details of the calibration). In the calibration experiment, the non-words were organized into 84 triads, each consisting of a Target (i.e., *Foka*), and two Test stimuli, such that Test B (i.e., *Fuka*) always shared more phonetic features with the Target than Test A (i.e., *Fema*). Children were asked to determine which of the Test stimuli sounded more like the Target.

Proportions of stimuli selection were converted into phonological similarity ratios. A total of 24 triads were selected to represent three levels of phonological similarity (see Table 1C for the list of selected triads). There were 8 triads at Ratio = 1 (Tests A and B sound equally similar to the Target), 8 triads at Ratio = 0.42 (Test A sounds somewhat less similar to the Target than Test B), and 8 triads at Ratio = 0.18 (Test A sounds markedly less similar to the Target than Test B). The set of triads with similarity ratio of 1 consisted of three identical items, and therefore was not calibrated. Note that unlike the visual triads, in which it was Test A that was equally or more similar to the Target, in the auditory triads it was Test B that was equally or more similar to the Target. As mentioned above, the non-identical labels were discriminable by young children.

The experiment had a 2 (Visual Similarity: Ratio = 1 vs. Ratio = 9) by 3 (Phonological Similarity: Ratio = 1, Ratio = 0.42, and Ratio = 0.18) mixed design, with visual similarity as a repeated measure. Participants were presented with 8 induction trials and 4 filler trials, with the order of trials randomized for each participant. Half of the trials were at Visual Similarity Ratio = 1 and another half were at Visual Similarity Ratio = 9. On each trial, participants performed an induction task, in which each of the Test items was described as having a particular quasi-biological property (e.g., long bones vs. short bones), and participants were asked to decide whether the Target item had long or short bones. Recall that in all trials, Test A had the same or greater appearance similarity to the Target than Test B, whereas Test B had the same or greater label similarity to the Target.

Procedure

Participants were interviewed individually in their childcare centers, with pictures presented on the screen of a portable computer, and labels presented through Sony-MDR-CD770 headphones. Participants were told that they would play a game, in which they would be shown pictures of animals, and the names of the animals would be played by the computer. The computer presentation of labels intended to preserve the calibrated similarity ratios in label triads. Participants were then presented with the induction trials.

Results and Discussion

Predicted and observed proportions of B-choices (i.e., induction from Test B to the Target) across the conditions are presented in Figure 6A and 6B, respectively. These observed proportions were subjected to a two-way mixed ANOVA with phonological similarity as a between-subjects factor and visual similarity as a repeated measure. There were two significant main effects, with no significant interaction ($p > .1$). The main effect of visual similarity, $F(1, 64) = 46.51$, $p < .0001$, $\eta_p^2 = .42$, replicated previous findings: as similarity of Test B to the Target decreased, participants were less likely to induce properties from Test B to the Target. More importantly, there was a main effect of phonological similarity: as phonological similarity of Test B labels to the Target labels increased, participants were more likely to induce properties from Test B to the Target, $F(2, 64) = 19.72$, $p < .0001$, $\eta_p^2 = .38$. Post-hoc Tukey tests revealed significant differences in performance among all three levels of phonological similarity, all ps $< .05$.

These results confirmed predictions of SINC, with both main effects being observed. Furthermore, a comparison of predicted and observed values indicates that SINC accounts for approximately 84% of the observed variance. Therefore, as predicted by SINC, phonological similarity of labels contributed to young children's induction.

Despite the overall accuracy, the model somewhat under-predicted the proportion of label-based responding in those conditions in which the two sources of information are in conflict (i.e., the visual ratio = 9, phonological ratio = 0.42 condition and the visual ratio = 9, phonological ratio = 0.18 condition). This is because our intention was to capture only quantitative tendencies in the data, and we did not include any free parameters in the current version of the model. However, attentional weights of labels are higher than these of visual features (see Sloutsky & Fisher, 2004), and when these differences are taken into account, predictions of the model become substantially more accurate.

GENERAL DISCUSSION

Several important findings stem from the reported experiments. Experiment 1 indicated that unlike adults, young children exhibit a phonological similarity effect: they extend phonologically similar labels to perceptually similar items. Results of the follow-up experiments with phonologically different novel labels (Experiment 2) and with phonologically similar familiar labels (Experiment 4) demonstrate that the obtained findings stem from the effects of phonological similarity of labels rather than from children's tendency to match pictures or to treat similar labels as tokens of the same words. Differential effects of phonological similarity on lexical extension in children and adults (Experiment 1) support previous findings suggesting that whereas young children are likely to treat labels as features of entities, adults may treat labels as symbols denoting category membership (Sloutsky & Lo, 1999; Sloutsky & Fisher, 2004; Yamauchi & Markman, 2000). Overall, Experiments 1-4 generated novel findings indicating that phonological similarity of labels contributes to early (but not to mature) generalization. These findings support the idea that different mechanisms may underlie early and mature induction, with early induction being similarity-based and mature induction being category-based (cf. Fisher & Sloutsky, 2005; Sloutsky & Fisher, 2004).

Experiment 5 demonstrated that increasing discriminability of labels attenuates young children's tendency to extend these labels to identical items. This finding provides converging evidence that young children's induction is driven by phonological similarity of labels while further undermining the possibility that it was driven by the tendency to interpret phonologically similar labels as tokens of the same word. Finally, Experiment 6 indicated that labels make a quantitative rather than an all-or-none contribution to early induction. In particular, it was demonstrated that identical labels make a greater contribution to induction than phonologically similar labels, however similar labels make a greater contribution than phonologically different labels. These results support predictions of SINC that phonological similarity of labels affects label extension and inductive generalization.

Phonological Similarity Effect: What is the Mechanism?

Recall that according to SINC, words are features contributing to similarity, and it is possible that effects of words are continuous rather than categorical. One possibility is that early in development, under many conditions, auditory information (including speech sounds) overshadows (i.e., attenuates processing of) corresponding visual input (Napolitano & Sloutsky, 2004; Sloutsky & Napolitano, 2003). Specifically, when visual stimuli were accompanied by sounds, discrimination of visual stimuli (but not sounds) dropped compared to a unimodal baseline. As a result of auditory overshadowing, entities associated with the same label become more similar (Sloutsky & Lo, 1999; Sloutsky & Fisher, 2004), with similarity driving both categorization and induction. The labels-as-features notion of SINC can both predict and account for current results as well as for the overshadowing effects.

Could it be that the reported results stem from other factors? In particular, it is possible that young children consider phonologically similar words as phonological tokens of the same word. Another, related possibility is that the effects are driven by some top-down conceptual assumptions. For example, young children may assume that phonologically similar words refer to similar entities. Although neither possibility undermines the novelty of reported findings, each possibility has different theoretical implications. In what follows, we consider each of these possibilities in greater detail.

Recall that these alternative possibilities were directly tested in Experiments 4 and 5, and results suggested that these possibilities are unlikely. First, results of Experiment 4 indicate that the phonological similarity effect persisted even when highly familiar words were used (e.g., “house” and “mouse”): young children generalized phonologically similar words to visually similar items. This finding is particularly difficult to reconcile with the token interpretation given that even 24-month-olds ably detect even small variations in pronunciation, especially when the words are familiar (e.g., Bailey & Plunkett, 2002). Second, results of Experiment 5 indicate that discrimination training reduced phonological similarity of words and the phonological similarity effect in the experimental condition (where the trained words were used in a subsequent generalization task), but not in the control condition (where the trained words differ from those used in the generalization task). Finally, results of Experiment 1 indicate that, unlike young children, adults exhibit little evidence of the phonological similarity effect. While the above described mechanism underlying effects of labels and developmental changes in these effects predicts all these findings, it is not clear how any of the alternatives could coherently account for these findings.

Overall, the reported research presents novel findings pertaining to effects of phonological similarity in a variety of induction tasks. These findings support the prediction of SINC, further indicating that for young children labels are features rather than category markers, and, as discussed below, these results affect our understanding of the mechanisms underlying early induction.

Labels, Similarity, Induction, and Categorization

Recall that according to SINC, effects of words on similarity and thus on induction stem from low-level attentional mechanisms rather than from an understanding of conceptual importance of labels (Sloutsky & Fisher, 2004; Sloutsky & Fisher, 2005). According to this view, at least early in development, words are features affecting processing of corresponding visual input and these effects of words could be (at least partially) grounded in auditory overshadowing effects. Therefore, words could be affecting inductive generalization by overshadowing differences among entities sharing the same label, thus making them appear more similar. The labels-as-features notion of SINC can both predict and account for current results as well as for the overshadowing effects.

Alternatively, it is possible that labels are category markers affecting categorization and induction by communicating category information, with children realizing that members of the same category share multiple commonalities (e.g., Gelman & Coley, 1990; Gelman & Markman, 1986, see also Gelman, 2003, Waxman, 2003, for reviews). According to this view, (a) labels affect inductive generalization in non-trivial ways that cannot be explained by low-level accounts and (b) these effects of labels manifest themselves even in infancy.

For example, in one study (Ferry, Hespos, & Waxman, 2010), 3- to 4-month-old infants were presented with a categorization task, such that members of a to-be-learned category were accompanied either by a common label or by a common tone. Results indicated that participants were more likely to categorize in the former than in the latter condition. In another study it was shown that lexical extension in 14- to 18-month-olds is more likely when words were presented as count nouns than when they are presented as adjectives (Booth & Waxman, 2009). Although these effects could be interpreted as evidence of labels being more than just features, we believe that these effects could be also explained by a low-level attentional mechanism. In particular, if familiar auditory stimuli are processed faster and more efficiently than novel stimuli, then (at least in infancy paradigms) more familiar auditory stimuli may have smaller interference effects than less familiar auditory input. As a result, less familiar tones may interfere with categorization to a greater degree than relatively more familiar speech sounds. Similarly, according to the lexical norms of the

MacArthur-Bates Communicative Development Inventories (Dale & Fenson, 1996), out of 384 words potentially comprehended by infants at 16 months of age, less than 10% are adjectives and more than 60% are nouns. Therefore, it seems likely that nouns represent a more familiar input than adjectives; consequently, adjectives may interfere with categorization to a greater degree than nouns. While there is no supporting evidence for this possibility with regards to nouns and adjectives, Robinson and Sloutsky showed that both words and unfamiliar sounds interfered with categorization (2007b) and individuation (2008) when compared to a silent condition, and that for infants this interference was stronger for less familiar auditory input.

Another finding that may potentially undermine the idea of labels as features is that participants are more likely to rely on linguistic labels when inferring a biological property, such as the type of heart of an animal than when inferring a physical property, such as the size of the animal (e.g., Gelman & Markman, 1986). However, there is recent evidence (Sloutsky & Fisher, 2008) that addresses this issue directly. A key idea is that many stimulus properties inter-correlate, such that some clusters of properties co-occur with particular outcomes, and other clusters co-occur with different outcomes. In particular, count nouns are more likely to co-occur with stable properties than with highly variable properties. Learning of these correlations may result in differential allocation of attention to predictive and non-predictive stimulus dimensions (cf., Kersten, Goldstone & Schaffert, 1998; Kruschke, 1992; Nosofsky, 1986). Therefore, if labels correlate with stable properties and do not correlate with variable properties, then participants should learn to attend to labels in the former, but not in the latter context. Given that size is a variable property (it varies between individuals and, for living things, size varies within individuals) and given that children have experience with this variability (i.e., a small puppy grows into a large dog), there is little surprise that children do not rely on labels in the context of size. Findings reported in Sloutsky and

Fisher (2008) support these notions pointing to successful attentional learning in 4- to 5-year-old children, who learned to rely differentially on the same predictor across different contexts, while exhibiting little or no awareness that learning had occurred.

The notion that early in development induction, categorization, and naming are similarity-based processes is able to account for a variety of other seemingly disjoint effects. For example, Markman and Hutchinson (1984) found that in the absence of a label, young children group things thematically (e.g., a police car and a policeman), whereas when the police car is named *a Dax* and the child is asked to select another *Dax*, they selected a passenger car. There is also a set of findings that toddlers and young children tend to interpret novel words as referring to novel items (Markman & Wachtel, 1988). The effect however reduces markedly or disappears if the novel word is phonologically similar to a familiar word: children would extend a novel word *Japple* to an apple rather than to a novel object (Merriman & Schuster, 1991).

These findings have been treated as stemming from different sources, with the first tendency being driven by the "taxonomic bias", which reflects early understanding that words refer to categories, the second tendency being driven by the "mutual exclusivity" assumption in young children (see, Markman, 1989, for a review of both the "taxonomic bias" and the "mutual exclusivity" assumption), and the third tendency being driven by children's interpretation of phonologically similar words as variants of the same word.

Current results suggest that all three effects may stem from the same tendency to extend words in a similarity-based manner (cf., Gentner, 1978; Landau, Smith, & Jones, 1998). While similarity-based induction and categorization give a straightforward explanation to Markman and Hutchinson's (1984) findings, the latter two effects require further consideration. Note that in the Markman and Wachtel (1988) studies, the novel label differed from that of the known entity, with the label dissimilarity possibly blocking the extension of this dissimilar label to the known entity (similar to results of Experiment 2 presented here). In the Merriman and Schuster (1991) studies the introduced label was similar to that of the known entity, which may promote the extension of this new label to the known entity (similar to results of Experiments 1, 3, and 5 presented here). The following hypothetical scenarios could illustrate this reasoning. First, as in Markman and Wachtel, a child is presented with an apple-like object and a novel object and asked to point to the *Dax*. Second, as in Merriman and Schuster, a child is presented with an apple-like object and a novel object and asked to point to the *Japple*. Finally, in a thought experiment, a child is presented with an apple-like object and a novel object and asked to point to the *Apple*. Based on Markman and Wachtel (1988) findings, we can expect that few children would select the apple-like object in the *Dax* condition, whereas based on Merriman and Schuster (1991), we should expect significantly more children to select the apple-like object in the *Japple* condition. Assuming that in the *Apple* condition children would overwhelmingly point to the apple-like object, the three scenarios would reveal the same trend as found in current research: very similar labels support induction more strongly than somewhat similar labels, whereas somewhat similar labels support induction more strongly than different labels.

Conclusion

Overall, the reported results indicate that phonological similarity of words contributes to early induction, driving inductive generalization across a variety of induction tasks. These findings shed light on how words affect young children's induction: they indicate that children may consider words as features contributing to the overall similarity of compared entities.

References

Bailey, T. M, & Plunkett, K. (2002). Phonological specificity in early words. *Cognitive Development*, *17*, 1265–1282.

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

Dale, P.S. & Fenson, L. (1996). Lexical development norms for young children. *Behavior Research Methods, Instruments, & Computers*, *28*, 125-127.

Dewhurst, S. A. & Robinson, C. A. (2004). False memories in children: Evidence for a shift from phonological to semantic associations. *Psychological Science*, 15, 782-786.

Ferry, A., Hespos, S.J., & Waxman, S. (2010). Language facilitates category formation in 3-month-old infants. *Child Development*, *81*, 472-479.

Fisher, A. V. & Sloutsky, V. M. (2005). When induction meets memory: Evidence for gradual transition from similarity-based to category-based induction. *Child Development*, *76*, 583-597.

Gelman, S. A. (2003). *The essential child.* New York: Oxford University Press.

Gelman, S. A., & Coley, J. (1991). Language and categorization: The acquisition of natural kind terms. In S. A. Gelman, S. & J. P. Byrnes (Eds.), *Perspectives on language and thought: Interrelations in development* (pp. 146-196). New York: Cambridge University Press.

Gelman, S. A., & Markman, E. (1986). Categories and induction in young children. *Cognition*, *23*, 183-209.

Gentner, D. (1978). A study of early word meaning: What looks like a jiggy but acts like a zimbo? *Papers and Reports on Child Language Development*, *15*, 1-6.

Jusczyk, P. W. (1997). *The Discovery of Spoken Language*. Cambridge, MA: MIT Press.

Keil, F. C., Smith, W. C., Simons, D. J., & Levin, D. T. (1998). Two dogmas of conceptual empiricism: Implications for hybrid models of the structure of knowledge. *Cognition*, *65*, 103-135.

Kersten, A.W., Goldstone, R.L., & Schaffert, A. (1998). Two competing attentional mechanisms in category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 1437-1458.

Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review*, *99*, 22-44.

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General*, *115*, 39-57.

Landau, B., Smith, L. B., & Jones, S. (1998). Object shape, object function, and object name. *Journal of Memory & Language*, 38, 1-27.

Markman, E. M. & Hutchinson, J. E. (1984). Children's sensitivity to constraints on word meaning: Taxonomic versus thematic relations. *Cognitive Psychology, 16*, 1–27.

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

Merriman, W. E., & Schuster, J. M. (1991). Young children's disambiguation of object name reference. *Child Development*, *62*, 1288–1301.

Murphy G. L. (2002). *The big book of concepts*. Cambridge, MA: MIT Press.

Napolitano, A. C., & Sloutsky, V. M. (2004). Is a picture worth a thousand words? The flexible nature of modality dominance in young children. *Child Development*, *75*, 1850-1870.

Robinson, C.W., & Sloutsky, V.M. (2004). Auditory dominance and its change in the course of development. *Child Development*, *75*, 1387-1401.

Robinson, C. W., & Sloutsky, V. M. (2007a). Visual processing speed: Effects of auditory input on visual processing. *Developmental Science*, *10*, 734–740.

Robinson, C. W., & Sloutsky, V. M. (2007b). Linguistic labels and categorization in infancy: Do labels facilitate or hinder? *Infancy*, *11*, 233-253.

Robinson, C. W., & Sloutsky, V. M. (2008). Effects of auditory input in individuation tasks. *Developmental Science*, *11*, 869–881.

Sloutsky, V. M., & Fisher, A. V (2008). Attentional learning and flexible induction: How mundane mechanisms give rise to smart behaviors. *Child Development*, *79*, 639-651.

Sloutsky, V. M. & Fisher, A. V. (2005). Similarity, induction, naming, and categorization (SINC): Generalization or inductive reasoning? Response to Heit and Hayes. *Journal of Experimental Psychology: General*, *134*, 606-611.

Sloutsky, V. M. & Fisher, A. V. (2004). Induction and categorization in young children: A similarity-based model. *Journal of Experimental Psychology: General*, *133*, 166-188.

Sloutsky, V. M., & Lo, Y.-F. (1999). How much does a shared name make things similar? Part 1: Linguistic labels and the development of similarity judgment. *Developmental Psychology*, *35*, 1478-1492.

Sloutsky, V. M., Lo, Y.-F., & Fisher, A. V. (2001). How much does a shared name make things similar: Linguistic labels and the development of inductive inference. *Child Development*, *72*, 1695-1709.

Sloutsky, V. M., & Napolitano, A. (2003). Is a picture worth a thousand words? Preference for auditory modality in young children. *Child Development*, *74*, 822-833.

Swingley, D., & Aslin, R. N. (2007). Lexical competition in young children's word learning. *Cognitive Psychology*, *54*, 99–132.

Waxman, S. R. (2003). Links between object categorization and naming: Origins and emergence in human infants. In Rakison, D., & Oakes, L. M. (Eds.), *Early category and concept development: Making sense of the blooming, buzzing confusion* (pp. 213-241). New York: Oxford University Press.

Yamauchi, T., & Markman, A.B. (2000). Inference using categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 776-795.

Appendix

Details of calibration of auditory stimuli

Auditory stimuli were calibrated in an experiment with 103 young children. In the calibration experiment, the non-words were organized into 84 triads, each consisting of a Target (i.e., *Foka*), and two Test stimuli, such that Test B (i.e., *Fuka*) always shared more phonetic features with the Target than Test A (i.e., *Fema*). To reduce the possibility that Test B words would be interpreted as morphological or inflectional derivations of Target words, differences between Test B and the Target were made not to approximate such derivations in English. Unlike English, where derivations are realized by adding a morpheme to the word (e.g., Dog → Doggie, Car → Cars), Test B items were created by changing parts of the stem of Target words. The triads were broken into blocks, with each triad presented three times within each block in random order, and child participants were asked to determine which of the Test stimuli sounded more like the Target (see Table 1C, for examples).

Table 1A

Label pairs used in Experiments 1-5

Unfamiliar phonologically similar labels used in Experiments 1, 3, and 5*	Unfamiliar phonologically different labels used in Experiment 2	Familiar phonologically similar labels used in Experiment 4
fika / pika	ti – gi / zuma	mouse – house
sudu / su – gu**	ta – ni / sodu	bone – boat
tabi / da – bi	no – bi / sadu	box – fox
fima / fina	sotu / ku – ma	horse – horn
zatu / gatu	gitu / vu – ka	dream – drum
gama / guma	fida / zo – tu	bug- mug
foka / fuka	tabi / fo – ka	bed – bell
sanu / sadu	satu / kipa	flower - shower

* The subset of label pairs used in Experiment 5 is underlined.

** Hyphenation indicates a long vowel.

Table 1B

Percent of children's label extension by test item type in Experiments 1-4

	Test Item			
	Identical	Similar	Less Similar	Dissimilar
Experiment 1 (unfamiliar similar labels)	84	9	4	3
Experiment 2 (unfamiliar dissimilar labels)	28	26	25	21
Experiment 3 (unfamiliar similar labels and "identical" item removed	N/A	78	19	3
Experiment 4 (familiar similar labels)	65	10	12	13

Table 1C

Label triads used in Experiment 6

	Label Type		
Similarity Ratio of Labels	Target Label	Test A Label	Test B Label
Ratio = 1 (Identical Labels)	si-du	si-du	si-du
	gama	gama	gama
	foda	foda	foda
	kipa	kipa	kipa
	zitu	zitu	zitu
	satu	satu	satu
	no-bi	no-bi	no-bi
	zi-bi	zi-bi	zi-bi
Ratio= 0.42 (Phonologically Similar Labels)	tobi*	sinu	todi*
	kuma*	gu-ma	ku-ma*
	zatu*	su-gu	ga-tu*
	gama*	botu	guma*
	sanu*	nabi	sadu*
	fika*	fo-da	pika*
	sotu*	fima	so-bu*
	foka*	fema	fuka*
Ratio= 0.18 (Phonologically Different Labels)	ti-gi	zuma	ti-gi
	ta-ni	sodu	ta-ni
	no-bi	sadu	no-bi
	sotu	ku-ma	sotu
	gitu	vu-ka	gitu
	fida	zo-tu	fida
	tabi	fo-ka	tabi
	satu	kipa	satu

Note: “*” denote phonologically similar labels. Hyphenation indicates a long vowel.

Figure 1. A. Examples of picture sets used in Experiments 1-5.

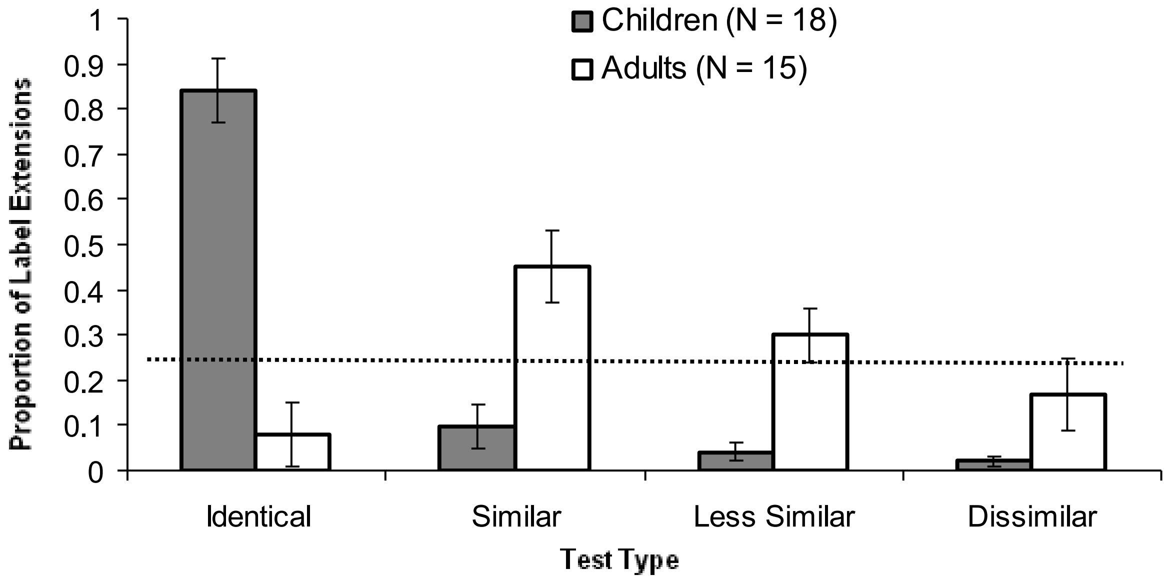


Figure 2. Proportion of label extensions by age and test item type and age in Experiment 1. Error bars represent the standard error of the mean, and the dashed line represents chance level.

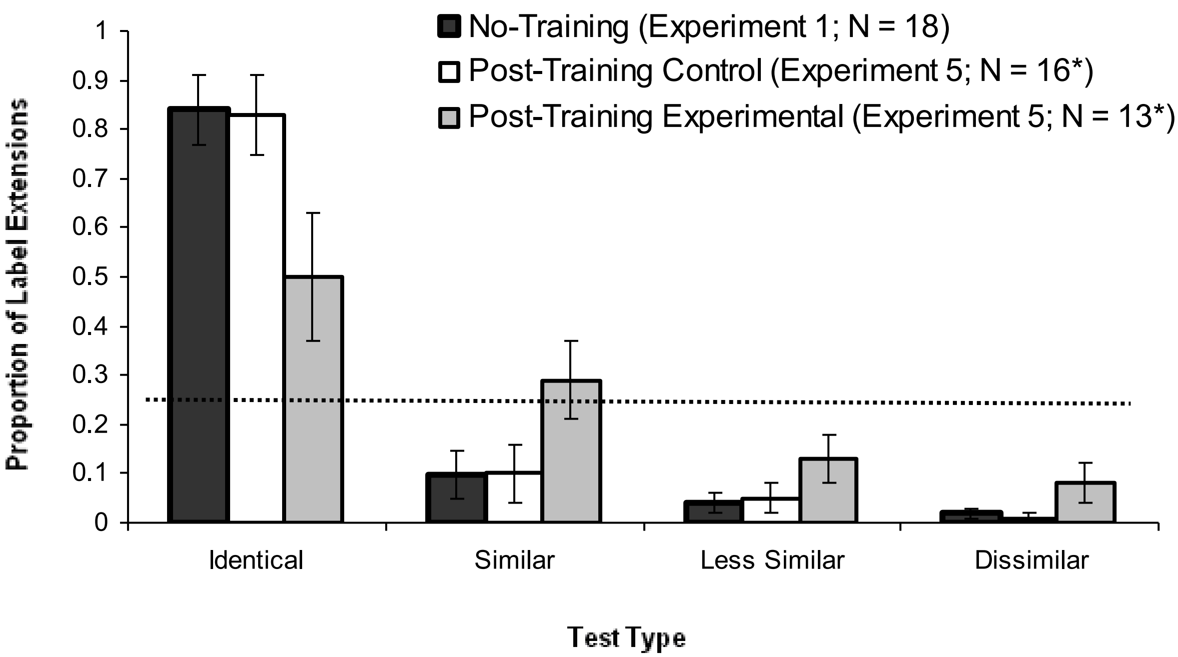
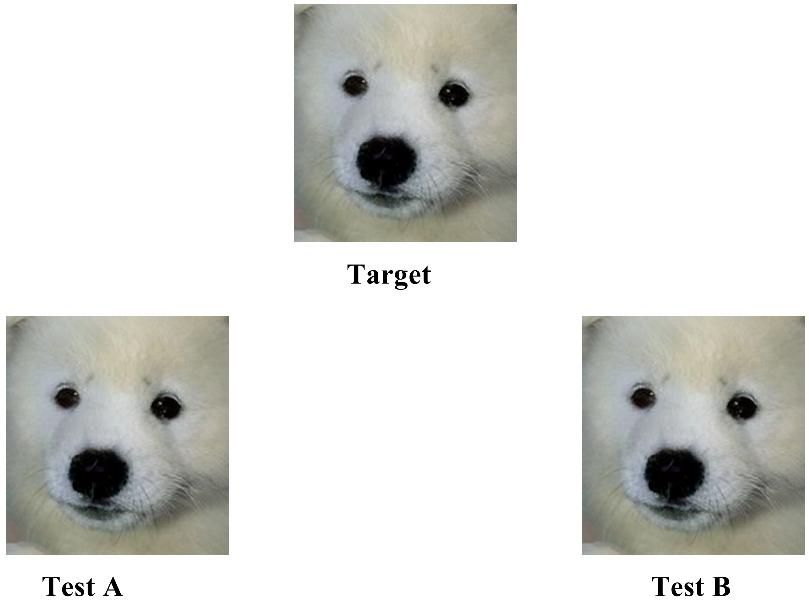


Figure 3. Proportions of label extensions by test item type and training condition in Experiment 5. Proportions of label extensions in Experiment 1 are included for comparison. Error bars represent the standard error of the mean, and the dashed line represents chance level.

* These Ns include only those children who reached 75% or above accuracy during training in Experiment 5.

(1) Appearances are non-predictive. Visual Similarity Ratio (A, Target)/(B, Target) = 1



(2) Appearances are fully predictive (Test A looks much more like the Target than Test B). Visual Similarity Ratio (A, Target)/(B, Target) = 9

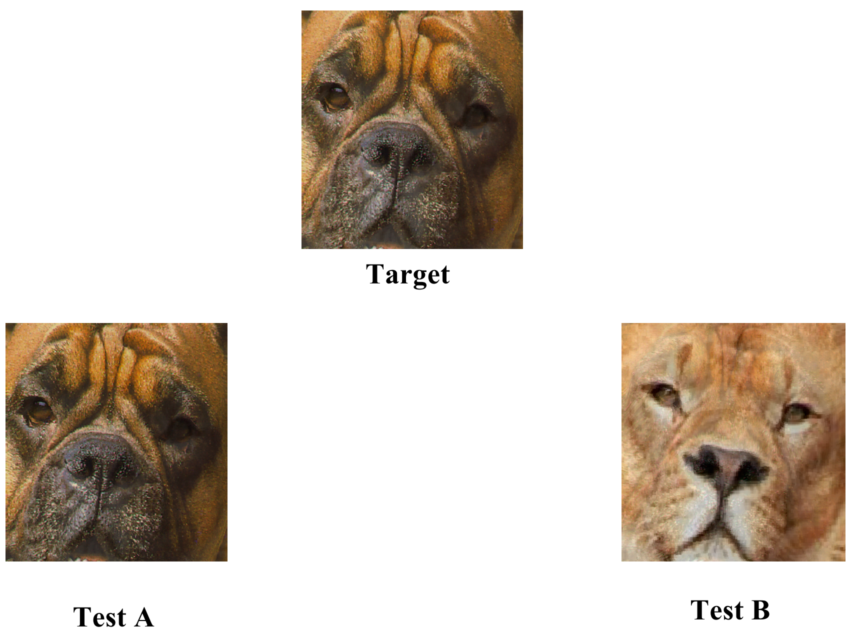
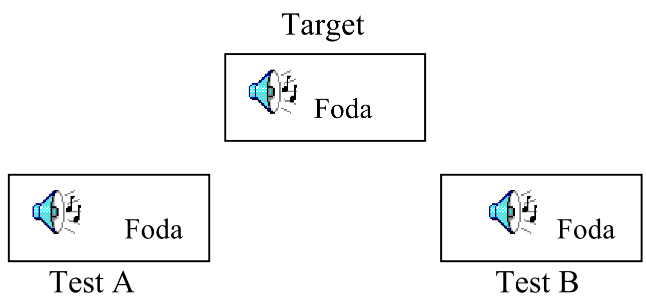
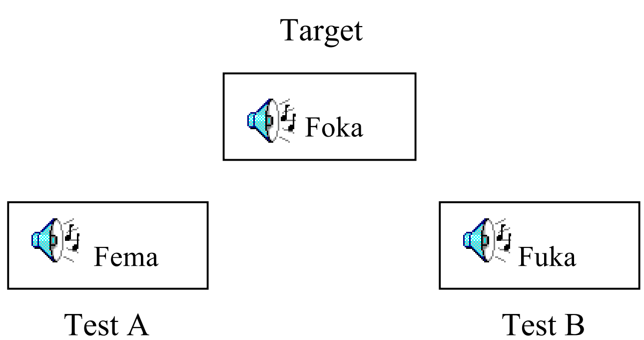


Figure 4. Examples of visual stimuli used in Experiment 6.

(1) Labels are non-predictive. Label Similarity Ratio (A, Target)/(B, Target) = 1



(2) Labels are partially predictive (Test B sounds somewhat more like the Target than Test A. Label Similarity Ratio (A, Target)/(B, Target) = 0.42



(3) Labels are fully predictive (Test B sounds much more like the Target than Test A. Label Similarity Ratio (A, Target)/(B, Target) = 0.18

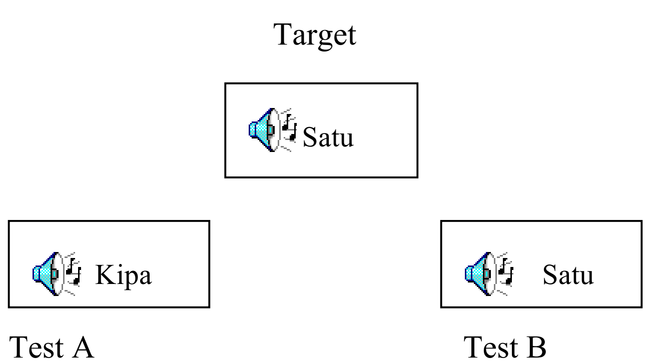
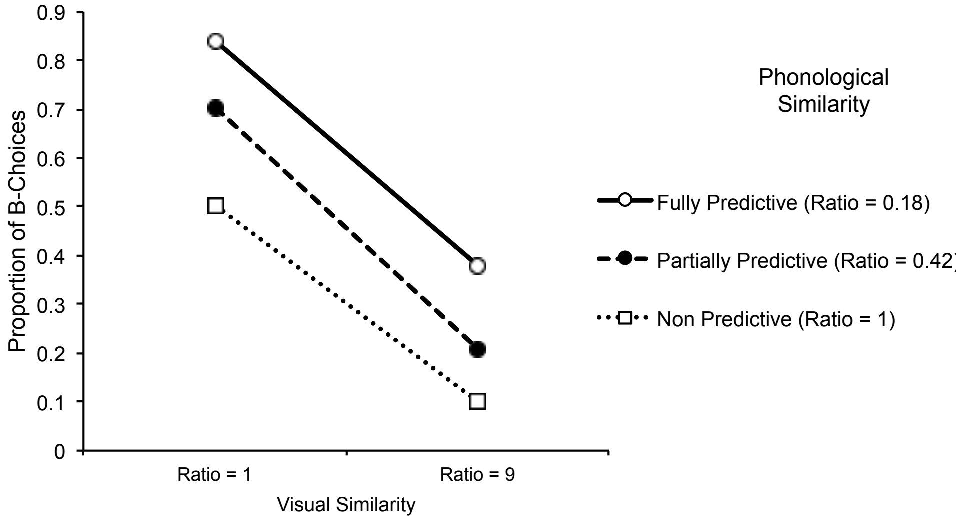


Figure 5. Examples of labels used in Experiment 6.

(A) Values predicted by SINC.



(B) Values observed in Experiment 6.

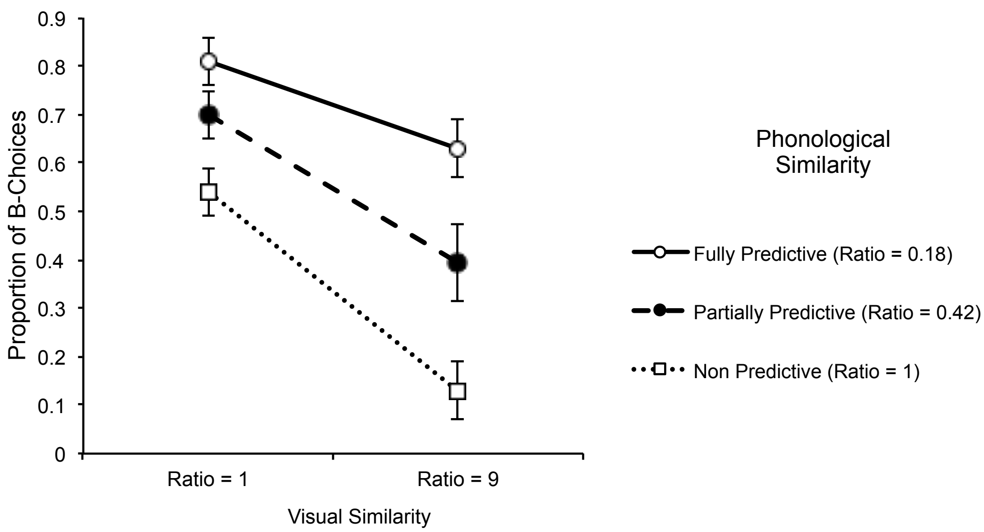


Figure 6. Proportions of B-choices across the levels of visual and phonological similarity. (A) Values predicted by SINC; (B) Values observed in Experiment 6.

Note: Phonological Similarity Ratio = 0.18: Test B's label is markedly more similar to the Target's label than Test A's label. Phonological Similarity Ratio = 0.42: Test B's label is somewhat more similar to the Target's label than Test A's label. Phonological Similarity Ratio = 1: Test B's label is as similar to the Target's label as Test A's label. Visual Similarity Ratio = 1: Test B is as similar to the Target as Test A. Visual Similarity Ratio = 9: Test B is markedly less similar to the Target than Test A.

Notes

[1] Using all eight labels pairs from Experiment 1 for discrimination training would make the experiment prohibitively long for young children, therefore, a random subset of 6 label pairs was used. Therefore, it is likely that effects of discrimination training observed in Experiment 5 are underestimated: greater effect of training could be observed if participants were given discrimination training on all eight label pairs used in Experiment 1.