Introduction

Words are often described as cognitive tools. On the basis of a label, a learner can extend information about a given exemplar to novel instances. For example, the fact that two animals are both labeled in the same way suggests that properties of one will be shared by the other. Like most other tools, we generally use those that our language and culture provide. The child relies on others to learn the appropriate ways to use words. This combination of social origin and individual application presents some characteristic challenges to the developing child. In particular, the sources of social information about labeling may vary in reliability. Given this, the characteristics of the people who provide a novel label may affect how the label is acquired and used. The current paper investigates whether the inductive role of a label depends on the reliability of the speaker who provides the label. Specifically, we ask whether children expect novel labels from informants who have been reliable in the past to have stronger inductive potential than labels from previously unreliable informants.

Gelman and her colleagues (e.g., Gelman, 2003; Gelman & Markman, 1986) showed that preschoolers make label-based inductive inferences. Children used a shared label, rather than a shared appearance, as the basis for property predictions. For instance, when two animals shared the same label, young children predicted that they shared the same unobservable property (e.g., having hollow bones) even though the two animals differed in appearance. Further, Jaswal and colleagues (e.g., Jaswal, 2004; Jaswal & Markman; 2007) showed that children rely upon label information over perceptual information when making inferences about a property of an animal. When an animal that

looked like a dog was called a cat, children made inferences (e.g., it drinks milk) based on the label (cat).

Although this evidence shows that children use labels provided by others to make inferences, a separate line of research reveals that children do not treat all labels equally. A growing body of research suggests that children selectively learn from others. Children around the age of four preferentially acquire novel labels from a reliable speaker who provided correct labels in the past over an unreliable speaker who provided incorrect labels in the past (e.g., Birch, Vauthier, & Bloom, 2008; Koenig, Clément, & Harris, 2004; Koenig & Harris, 2005). Children are also more likely to acquire labels from a speaker who expressed labels with confidence (Sabbagh & Baldwin, 2001) and from a speaker who knows the conventions for labeling (Diesendruck, Carmel, & Markson, 2010). Indeed, this selective trust guides children's subsequent learning from the same speakers even after a 1-week delay (Corriveau & Harris, 2009). This work has focused on the acquisition of labels - how a child decides what something is to be called. Thus, children are more likely to agree that a novel object is called a "dax" if a reliable, confident, conventional speaker provides the label. Our question is whether this sensitivity to speaker reliability extends to future inferences from the new label.

Thus, the current paper brings two research domains together by examining whether children's selective trust in others' testimony guides their inductive inferences. Will children infer that two objects that share a label but differ in appearance share novel properties when the label is provided by a reliable speaker but not when the label is provided by an unreliable speaker? Does the past reliability of the informant influence children's expectations that two objects both called "dax" will share novel properties?

Below we review these two lines of research in more detail before describing the way that we sought to bring them together.

Previous work by Jaswal and his colleagues has documented the impact of the speaker on young children's willingness to make inductive inferences with familiar categories (Jaswal, 2004, 2006; Jaswal & Malone, 2007). For example, Jaswal and Malone (2007) demonstrated that the confidence of a speaker influences children's label-based inferences. They presented 3-year-olds with a hybrid picture in which appearance indicated membership of one category membership whereas a label indicated membership of another. For instance, children were shown a picture of an object that looked like a key but was labeled as a spoon. Then, children were asked to infer the function of the object (e.g., "starts the car" vs. "used to eat cereal from the bowl"). Critically, when the speaker expressed uncertainty about her claim (e.g., "*I think* this is a spoon") along with other behavioral cues such as furrowing the brow and hesitancy, children's inference were less likely to be based on the provided label than when a speaker provided the label without such uncertainty cues (e.g., "This is a spoon").

However, despite a growing body of work on children's selective learning from reliable speakers (e.g., Birch et al., 2008; Koenig & Harris, 2005), very little prior work has investigated the effect of speaker reliability on children's novel inductive inferences. For example, Jaswal and Malone's study was designed to look at whether children were willing to accept *familiar labels* from speakers who varied in confidence when the labels that they provided were unexpected and conflicted with children's own appearance-based expectations. The child had to decide whether the object was a key or a spoon and having done so, could infer the properties normally associated with that category. The child did

not learn any new properties of keys or spoons. By contrast, we asked how children use *novel labels* from reliable as compared to unreliable speakers to guide their future learning. Additionally, although novel artifact categories have been used to study children's inductive inferences (e.g., Gelman, 1988; Graham, Kilbreath, & Welder, 2004) such inferences have not been examined in the context of variation in speaker reliability. Hence, we asked how far children use novel labels from reliable as compared to unreliable speakers to guide their subsequent inferences. More specifically, we asked how children learn the new and non-obvious properties of a category and extend those properties to other novel objects belonging to the same category. Thus, the current paper extends prior work by investigating whether speaker reliability influences children's tendency to use labels to make inferences about the non-obvious properties of novel objects.

Insofar as previous research has shown that the speaker of a label influences children's inductive inferences concerning familiar categories (Jaswal, 2004, 2006; Jaswal & Malone, 2007), comparable effects may occur for novel artifact categories. On the other hand, given that inductive potential is less robust for artifact categories as compared to natural kind categories (see, Gelman, 1988), it is an open question whether a speaker would have *any* effect on children's inductive inferences with novel artifact categories. Suggestively, in a recent study by Sobel and Corriveau (2010), children endorsed novel object labels from an informant who had reliably predicted non-obvious internal properties of objects. In their study, two informants differed in their expert knowledge about the internal properties of objects. One informant provided accurate predictions about the redness of a novel object's inside but was ignorant ('I don't know')

about its greenness whereas the other informant was accurate about the greenness of an object's inside but was ignorant about its redness. By 4 years of age, children endorsed labels from an appropriate informant depending on the relevant property of an object. For example, when presented with an object that was red inside, children endorsed a label from the red rather than the green expert. Interestingly, informant's differential knowledge about the external properties of the object (red vs. green stickers on the back) did not influence children's selective endorsement of the labels offered by two informants. This study showed that by 4 years of age, children treated an informant's reliability concerning non-obvious internal properties as an index of his/her knowledge about the label for a given novel object. The current study asks whether a speaker's reliability in labeling guides children's inductive inferences about internal, non-obvious properties.

Our inductive inference task closely followed Gelman's triad method (see Gelman & Markman, 1986). Thus, one target object and two test items were presented. One of the test items was similar in appearance to the target but received a different label. The other test item shared the same label as the target but was dissimilar in appearance. Children were asked to predict which of the two test items also had a property that was ascribed to the target item. In the current study, the object that was similar to the target was highly similar whereas the object that shared a label was markedly dissimilar. Hence, the shared label was in competition with a shared appearance.

In order to vary the perceived reliability of the speaker, we used a technique that has been successful in previous studies (Koenig et al., 2004). Children first received a reliability establishment phase in which one speaker named familiar objects accurately

whereas the other speaker named the same familiar objects inaccurately. Following this establishment phase, children proceeded to a test phase, involving one of the two speakers. After a target object was named and attributed a hidden property (e.g., being magnetic) by the experimenter, either the reliable speaker or the unreliable speaker established a conflict between appearance and labels by naming the similar test object with a different label from the experimenter and naming the dissimilar test object with the same label. Children were then asked to say which of the two test objects possessed the hidden property (e.g., "which one of these is magnetic?"). It was anticipated that selection of the object with the same label as the target (despite its conflicting appearance) would be more frequent among children who were supplied that label by a reliable rather than an unreliable speaker.

As a further check on the impact of labeling by reliable as compared to unreliable speakers, a control group of children was tested using the same materials but none of the objects was labeled and children received no information about speaker reliability. It was anticipated that children in this no-label condition would make appearance-based inferences. Effectively, then, their pattern of responding should be similar to that produced by children experiencing an unreliable speaker.

In a second inference task, we asked how far children's experience of a conflict between the appearance of the two choice objects and the labels supplied for them by differentially reliable speakers would influence their subsequent predictions about a hidden property *in the absence of labels*. Children were presented with two test objects, one, similar to the target, and the other, completely novel. Children were then asked to select the test object that did not have the hidden property as the target (e.g., "Which one

of these is not magnetic?"). Given that there were no labels provided for the choice objects in this second inference task, the shared appearance was the only available guide for making an inference. However, the label-appearance conflict established in the first inference task was expected to have a greater impact on children's subsequent inferences with a different set of objects when that conflict had been provoked by the reliable as opposed to the unreliable speaker. Specifically, if the reliable speaker had presented children with the label-appearance conflict, they should be less confident in making an appearance-based inference. By contrast, if the unreliable speaker had presented children with the label-appearance conflict, they should retain confidence in making an appearance-based inference. Therefore, as compared to children in the reliable condition, children in the unreliable condition were expected to be guided by the relative similarity of the choice objects to the target and to select the markedly different object as the one that would lack the hidden property.

In previous research, the reliability of the speaker has been mainly studied with children ranging from 3 to 5 years. Similarly, label-based inferences have mostly been studied with children ranging from 3 to 5 years. Accordingly, we tested 3- and 5-year-olds.

In summary, the current paper asks whether the inductive role of a label depends on the reliability of the speaker who provides it. To what extent does a speaker's reliability influence children's willingness to infer a non-obvious property from the label supplied by that speaker and how far does speaker reliability affect children's subsequent learning about a different set of novel objects?

Method

Participants

Thirty-two (18 girls) 3-year-olds (M = 3.7, range = 3.1 – 3.9) and twenty-nine (18 girls) 5-year-olds (M = 5.4, range = 5.0 – 5.9) participated in the study. Sixteen (M = 4.6, range = 3.2 – 5.9), seventeen (M = 4.6, range = 3.7 – 5.5), and twenty-eight (M = 4.7, range = 3.1 – 5.7) children were assigned to the reliable speaker, unreliable speaker, and no label condition respectively. Most children were Caucasian and from middle or upper-middle class families. They were recruited from local daycares and preschools. For all children, English was their native language.

Design and procedure

The procedure included an initial establishment phase (for children in the reliable and unreliable speaker conditions) and a subsequent test phase (for children in all three conditions). In the reliable speaker condition, the speaker provided correct labels for the initial pair of objects in the establishment phase. In the unreliable speaker condition, the speaker provided incorrect labels for the initial pair of objects in the establishment phase. In the no label condition, the initial pair of objects was not given labels at all in the establishment phase. Every child received two blocks. Each block consisted of establishment phase (in the reliable and the unreliable speaker conditions) and a testing phase. There were two trials in each test phase. Details of the establishment phase and subsequent testing phase involving two pairs of objects are given below.

Speaker reliability establishment phase. During this phase, two puppets were introduced, Lion and Rhino. One puppet was established as reliable by providing correct labels for four familiar objects and the other puppet was established as unreliable by providing incorrect labels for the same objects. For instance, encountering a key, the

reliable puppet said, “This is a key,” and the unreliable puppet said, “This is a pencil.” Lion always spoke first, but for half of the children, Lion was reliable, and for the other half, Rhino was reliable. The location of the puppet (to the left or right of the child) was randomized. After each establishment phase, the experimenter posed a manipulation check question about both the reliable and unreliable puppets, “Do you think he is good at naming things?” and corrected children who provided wrong answers. All children responded correctly, except for one 3-year-old, who gave wrong responses twice. Each of the two blocks began with an establishment phase. That is, the speakers’ reliability status was re-established to ensure that the difference between the speakers remained salient for children. Therefore, across the establishment phase of the two blocks children saw the same two speakers, one providing correct labels and the other providing incorrect labels for a total of 8 different familiar objects.

Children in the no-label condition did not receive the speaker establishment phase. They proceeded directly to the testing phase.

Testing phase. The testing phase consisted of two inference tasks. First, the experimenter presented a novel target object and a novel label for the target, “Look at this! Isn’t this nice? This is a fep.” A non-obvious property of the object was then introduced by the experimenter, “I will tell you something about this. This is magnetic. I will show you.” Then, the experimenter demonstrated the property. The four properties used for this study were being magnetic, having a spring inside, lighting up, and rattling when shaken. Property order was randomized across children. Children were not permitted to interact with the test items until the study was completed.

After the demonstration of the non-obvious property, children received the first inference task. The experimenter presented two test items. One of the test items was similar in appearance to the target and the other test item was dissimilar in appearance to the target. The experimenter said, "I have these two things. Let's show these to one of these guys." For children in the reliable and unreliable speakers conditions, the relevant puppet provided labels for each of the two test items. The other puppet was removed from the child's view. For the dissimilar test item, the puppet provided the same label as the target object (henceforth same label/different appearance object), "This is a fep." For the similar test item, the puppet provided a different label (henceforth same appearance/different label object), "This is a tog." The experimenter then asked children to infer which of the two test items also had the same property as the target, for instance, "Which one of these is magnetic?" We expected that children's property prediction would be more often based on a shared label rather than on a shared appearance when the labels were provided by a reliable as opposed to an unreliable speaker. Thus, children were expected to select same label/different appearance objects more frequently in the reliable speaker condition than in the unreliable speaker condition. Note that the target remained in front of the child throughout the first inference task. The novel labels used were *fep (tog), yem (bem), lima (mido), wug (dax).*

The procedure adopted for children in the no label condition was largely the same as above except for the following changes: first, as noted earlier, there was no speaker reliability establishment phase. Children were simply introduced to two puppets. Additionally, the experimenter demonstrated a non-obvious property without the use of a label. Then, in the first inference task, instead of providing labels, a puppet commented

on the presented test objects, saying, “Look at these! These are very nice!” The other puppet was removed from the child’s view. For half of the children, Lion commented on the objects, and for the other half, Rhino commented on the objects. In the absence of any labels and any clues to speaker reliability, it was expected that children would select the same appearance rather than the different appearance choice.

After the target object and the two test objects had been removed, children received the second inference task. The experimenter provided two new test items. One was similar to the original target object (and therefore to the same appearance/different label object) whereas the other did not resemble any of the objects in the first inference task (see Appendix A for an example of object stimuli). The former will be referred to as the highly similar test object and the latter as the completely novel test object. Critically, no labels were provided for the objects in this second inference task. Children were asked to predict which of the two test items *did not* share the same property as the target. For example, the experimenter asked, “Can you give me the one that is not magnetic?” Because no label was provided for the objects in this second inference task, the only available information was similarity of appearance to the target object. Note that children in the reliable and unreliable speaker conditions had encountered a conflict between the label and appearance of each choice object in the first inference task. They were confronted with two dissimilar objects (the target and the dissimilar choice object) sharing a label and two similar objects (the target and the similar choice object) not sharing a label. However, this conflict was only likely to matter to children in the reliable speaker condition. It was expected to undermine their confidence in appearance-based inferences. By contrast, the conflict was less likely to affect children in the unreliable

speaker condition. They should continue to make appearance-based inferences. More specifically, they should conclude that a highly similar object would share the hidden property with the target whereas a completely novel object would not. Thus, children were expected to select completely novel objects more often in the unreliable speaker condition than in the reliable speaker condition. Note that, as described above, the second block began with another reliability establishment phase.

For children in the no label condition, it was expected that they would make appearance-based inferences. Thus, they would select the completely novel choice rather the highly similar choice when asked to identify the choice lacking the hidden property. After both blocks were completed, children were asked to name the familiar objects used in the establishment phase to confirm that they knew the names. All the children answered correctly. Upon the completion of the testing, when children were given a chance to interact with the novel objects they were asked if they have seen them before. All children denied having seen them before. Note that individual children were tested in a separate room of their school.

Results

Children's selection of a same label/different appearance object in the first inference task was coded as a 1 and their selection of a same appearance/different label object was coded as 0 for both the reliable and the unreliable speaker conditions. Children's selection of dissimilar looking objects (equivalent to the same label/different appearance objects) was also coded as the dependent measure for the no label condition. Figure 1 depicts the frequency with which children selected same label/different appearance (dissimilar looking objects in the case of the no label condition) objects as a

function of age and condition. Because there were 2 blocks children could receive a score ranging from 0 – 2 in each condition. Inspection of Figure 1 shows that children in the reliable condition were more likely to make label-based inferences than children in the unreliable and no label conditions.

To check this conclusion, inference scores were analyzed with a 2-way 2 (Age group: 3-year-olds and 5-year-olds) X 3 (Condition: a reliable speaker, an unreliable speaker, no label) ANOVA with Age and Condition as between-subject factors. This analysis confirmed the main effect of Condition, $F(2, 55) = 4.47$, $p < .05$, $h^2 = .14$. The main effect of Age was not significant, $F(1, 55) = .08$, n.s., nor was the interaction of Age and Condition, $F(2, 55) = .32$, n.s.. Given that scores ranged only from 0-2, we re-ran this analysis, using proportional scores and an arcsine transformation. This yielded an equivalent pattern of results. Thus, only the main effect of Condition reached significance, $F(2, 55) = 4.53$, $p < .05$, $h^2 = .14$.

Post-hoc analyses (Tukey's HSD) were carried out on the main effect of Condition. Children in the reliable condition selected same label/different appearance objects significantly more often than did children in the unreliable condition or in the no label condition (both comparisons $p < .05$). There was no difference between the unreliable condition and the no label condition. An equivalent pattern of results emerged when the three conditions were compared via Mann-Whitney U tests. Children in the reliable condition chose same label/different appearance objects more often than children in the unreliable ($U = 81$, $p < .05$) and no label ($U = 136$, $p < .05$) condition whereas these latter two conditions did not differ from one another ($U = 232.5$, n.s.)

Note that although children were more likely to make label-based inferences in the reliable speaker condition, children's selection of the same label/different appearance objects was no greater than chance, t (15) = - 0.62, n.s. In the unreliable speaker condition and the no label condition children typically assumed that objects with the same appearance shared properties with the target. Thus, they chose the same label/different appearance objects significantly less frequently than expected by chance, Unreliable speaker condition, t (17) = -6.20 p <.001, No label condition, t (27) = -5.87, p <.001. By implication, when children heard a label from a reliable speaker they were likely to suppress their very systematic default tendency to make an appearance-based inference (as displayed in the unreliable and no label conditions).

For the second inference task, children's selection of the completely novel objects (and the equivalent objects in the no label condition) was coded as the dependent measure. Figure 2 shows children's mean scores as a function of age and condition. Inspection of Figure 2 shows that children in the reliable condition chose the completely novel object less often than children in the unreliable and no label conditions. A 2 X 3 ANOVA of Age group (3-year-olds vs. 5-year-olds) X Condition (reliable speaker, unreliable speaker and no label condition) was carried out with Age and Condition as between-subject factors. This confirmed the main effect of Condition, F (2, 55) = 4.89, p < .05, h^2 = .15. The main effect of Age was not significant, F (1, 55) = .01, n.s., nor was the interaction of Age X Condition, F (2, 55) = 1.57, n.s.. As with the earlier ANOVA, because scores ranged only from 0-2, we re-ran this analysis, using proportional scores and an arcsine transformation. Again, this yielded an equivalent pattern of results. Thus, only the main effect of Condition reached significance, F (2, 55) = 4.04, p < .05, h^2= .13.

Post-hoc analyses (Tukey's HSD) revealed a significant difference between the reliable condition and the unreliable condition and between the reliable condition and the no label condition (both comparisons $p < .05$). The unreliable condition was not significantly different from the no label condition. Again, a similar pattern emerged when the conditions were compared with Mann-Whitney U tests. Children in the reliable condition chose the completely novel object less often than children in the unreliable (U=76 $p < .05$) and no label (U= 144 $p < .05$) conditions but these latter two conditions did not differ from one another (U= 224 n.s.).

Consistent with predictions, when children were asked to indicate which of the two test items *did not* share the same property as the target they often avoided completely novel objects in the reliable condition. In contrast, the completely novel objects were frequently chosen by children in the unreliable and no label conditions.

Children selected completely novel objects more frequently than expected by chance both in the unreliable condition, $t(16) = 6.2$, $p < .001$, and in the no label condition, $t(27) = 4.38$, $p < .001$ whereas selection of completely novel objects in the reliable condition was no different from chance, $t(15) = 0$, n.s. Thus, children were more likely to suppress the strategy of selection by appearance when the label had been provided by a reliable speaker.

In order to examine children's overall performance in both inference tasks as well as individual patterns of performance, children were scored for the number of times (maximum = 4) that they made an appearance-based choice. Table 1 shows the proportion of children in each condition scoring 0, 1, 2, 3 or 4. Inspection of Table 1 confirms that the majority of children in the unreliable and no label conditions scored 4

whereas only a minority of children in the reliable condition did so. Indeed, a minority of children in the reliable condition scored 0, reflecting their systematic label-based choices. Mann Whitney U tests confirmed that the reliable condition differed from both the unreliable condition (U= 72.5. $p < .05$) and the no label condition (U = 132.5, $p < .05$), whereas the two latter conditions did not differ from one another (U = 237, n.s.). Thus, the results of these non-parametric tests were consistent with the pattern revealed by the two earlier ANOVAs.

Finally, we conducted Binomial tests to assess how far the performance of individual children, as revealed in Table 1, deviated from chance. These tests confirmed that a greater proportion of children made appearance-based choices on all four trials than would be expected by chance, both in the unreliable condition ($p< .001$) and in the no label condition ($p < .001$). By contrast, analysis of children in the reliable condition showed that not only did a greater proportion of children make appearance-based choices on all four trials than would be expected by chance ($p< .05$) but in addition a greater proportion of children avoided making appearance-based choices on all four trials (i.e. made label-based choices) than would be expected by chance ($p<.05$). Thus, performance in the reliable condition was non-random but divided between two different strategies. To further confirm our findings, another analysis was conducted to directly compare across the reliable and unreliable conditions. First, the distribution of children who consistently used appearance-based choice (those who scored 4) and children who did not was compared between the reliable and unreliable conditions. Similarly, the distribution of children who consistently used label-based choice (those who scored 0) and those who did not was compared between the reliable and unreliable conditions. Fisher Exact Tests

revealed no significant difference between the two conditions in terms of the distribution of children using the appearance based choice, n.s.. By contrast, the distribution of children using label-based choice was different between the two conditions, $p < .05$.

Discussion

The current study shows that the reliability status of the speaker who provides a label plays an important role when young children make inductive inferences about non-obvious properties. Consistent with prior work (e.g., Jaswal 2004, 2006; Jaswal & Malone, 2007), the findings show that not every label has the same inductive power. The *speaker* who provides the label has a moderating effect on children's inductive inferences. More specifically, in the current study, children were unlikely to assume that objects sharing labels would also share properties when an unreliable speaker as opposed to a reliable speaker had provided the labels. Moreover, children used speaker reliability to form some expectations about the relations between labels and appearances for novel objects. Children who had been told by a reliable rather than an unreliable speaker that two dissimilar objects shared the label in the first inference task were less prone to make appearance-based choices with a new set of objects in the absence of labels in the second inference task. That is, information from a reliable speaker led children to expect that appearance was a relatively poor guide to shared properties.

The precise role of shared labels in inductive inference has recently been debated in the literature (Gelman & Waxman, 2007, 2009; Sloutsky, 2009; Sloutsky, Kloos, Fisher, 2007a, 2007b; Waxman & Gelman, 2009). One proposal is that a label is just like any other perceptual feature of an object. On this view, shared labels make two objects more similar and the overall shared similarity underlies the extension of a given property

from one object to another (e.g., Sloutsky & Fisher, 2004; Sloutsky, Lo, & Fisher, 2001). An alternative proposal is that shared labels indicate other relevant properties because a label provides direct access to category membership and it is expectations about categories that guide property inferences (Gelman, 2003). How do our findings fit with these two proposals?

Our findings show that not every label supports the same inductive inferences. Instead, the source of a label matters, that is, how reliable a speaker was in her past labeling. If a label is just another feature of an object then it is unclear why the same label would be treated differently depending on a speaker's past reliability. On the other hand, the category-label account can accommodate the current findings. On this account, a label is not just like any other perceptual feature because it conveys a chain of knowledge, particularly kind-relevant information that is transmitted via communication. A label invites a child to accept a body of expert knowledge, in line with the "division of linguistic labor" proposed by Putnam (1975). In the child's mind, a label acts as a theory that identifies a kind-relevant property and guides its extensions. Consistent with this, Jaswal (2004) noted that children's label-based inferences are theory-like in the sense that children's implicit assumption guides the selection of not only which properties are relevant to categories but also which source is worth paying attention to (see also Gelman, 2003).

Category-label inferences depend both on the assumption that objects that share a label belong to the same kind category, and the assumption that speakers who provide labels have communicative intentions, beliefs, and knowledge. A label allows one access to theory-rich knowledge but this does not happen in a social vacuum. Given this, as a

learner, one does not want to learn and rely on any label encountered. One wants to learn and use labels that reflect the established wisdom of one's community.

Unlike previous studies involving dissimilar objects (e.g., Gelman & Markman, 1986), the dissimilar objects in the current study were maximally dissimilar from the target whereas the similar objects displayed a maximal overlap of perceptual features (see Appendix A for an example). Given this, it is noteworthy that some children in the reliable condition showed systematic label-based inferences. One possible objection to the category-label account is that participants often used perceptual similarity to predict novel features. Our conclusions are based on the relative significance of labels given reliable and unreliable speakers, not the absolute rates of label use. The central point is that children are more likely to use a label from a reliable speaker. The similarity-based account might propose that when a speaker proves reliable, a single shared feature (i.e., a label) is heavily weighted and can trump a maximal non-overlap of perceptual features, thereby rendering the objects similar. However, it is unclear how speaker reliability itself can be construed in terms of similarity. Perhaps, the similarity-based account could marry the computation of a person with that of labels to explain the finding of relative difference in children's use of labels between reliable and unreliable speakers.

Why did more children not make systematic label-based inferences? We suspect that the most plausible explanation lies in our choice of test stimuli. As noted, one of the two test items was maximally different from the target (despite the common label) whereas the other test item was maximally similar to the target (despite the absence of a common label). The strong pull of appearance was revealed by the overall pattern of performance in both the unreliable and no label conditions. Thus, if children were to be

guided by the reliable speaker, they needed to set aside the strong pull of common appearance.

The present study was designed to bring together two important lines of research. First, we know that children use labels as a basis for inductive inferences (e.g., Markman & Gelman, 1986). The present findings show that this effect is restricted to labels provided by reliable speakers. Children do not base inductive inferences on labels provided by unreliable speakers. Thus, exactly the same label delivered by a reliable versus an unreliable speaker confers differential inductive power. We also know that children selectively learn from others (e.g., Birch et al., 2008; Diesendruck et al., 2010; Koenig & Harris, 2005; Sabbagh & Baldwin, 2001). For this line of research, the current findings suggest that children's selective trust in a speaker's labeling goes beyond the moment when the label is encoded – it guides children's later inferences. Indeed, children continued to keep in mind what they had learned from the speakers when subsequently presented with a different set of novel objects in the second inference task. Thus, children do not simply absorb what they hear from other people. Instead, they selectively adopt, use, and integrate what others say in constructing their own conceptual understanding of the world. The current study suggests that the labels children hear from a reliable speaker provide them not just with a communicative tool but a conceptual tool to guide their inferences about hidden properties.

Although it is unlikely that children of this age have any deep knowledge about the properties in question (e.g., magnetism), it is possible that differential experience with a given property might lead children to rely more or less on speakers. However, even if children in the current study had prior experience with some of the properties, their

reliance on the speaker's labels in their predictions about other novel objects was influenced by the speaker's past accuracy. Future research should address how children's preexisting beliefs and conceptual systems interact with this selective trust. More specifically, how do individual differences in experience with a given property influence their reliance on informants? And how flexibly do children update their conception of others as either reliable or unreliable? What role do changes in children conception of others' reliability play in their conceptual understanding? By investigating these questions, we will have a better understanding of the interplay between children's selective trust and their conceptual understanding.

References

Birch, S. A. J., Vauthier, S. A., & Bloom, P. (2008). Three- and four-year-olds spontaneously use others' past performance to guide their learning. *Cognition, 107*, 1018-1034.

Corriveau, K., & Harris, P. L. (2009). Preschoolers continue to trust a more accurate informant 1 week after exposure to accuracy information. *Developmental Science, 12,* 188-193.

Diesendruck, G., Carmel, N., & Markson, L. (2010). Children's sensitivity to the conventionality of sources. *Child Development, 81,* 652-668.

Gelman, S.A. (1988). The development of induction within natural kind and artifact categories. *Cognitive Psychology, 20*, 69-95.

Gelman, S. A. (2003). *The essential child: Origins of essentialism in everyday thought.* London: Oxford University Press.

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

Gelman, S. A., & Waxman, S. R. (2007). Looking beyond looks: Comments on Sloutsky, Kloos, and Fisher (2007). *Psychological Science, 18*, 554-555.

Gelman, S. A. & Waxman, S. R. (2009). Taking development seriously: Theories cannot emerge from associations alone. *Trends in Cognitive Sciences, 13*, 332-333.

Graham, S.A., & Kilbreath, C. S., & Welder, A. N. (2004). Thirteen-month-olds rely on shared labels and shape similarity for inductive inferences. *Child Development, 75*, 409-427.

Jaswal, V.K. (2004). Don't believe everything you hear: Preschoolers' sensitivity to speaker intent in category induction. *Child Development, 75*, 1871-1885.

Jaswal, V. K. (2006). Preschoolers favor the creator's label when reasoning about an artifact's function. *Cognition, 99,* B83-B92.

Jaswal, V.K., & Malone, L. S. (2007). Turning believers into skeptics: 3-year-olds sensitivity to cues to speaker credibility. *Journal of Cognition and Development, 8*, 263-283.

Jaswal, V. K., & Markman, E. M. (2007). Looks aren't everything: 24-mont-olds' willingness to accept unexpected labels. *Journal of Cognition and Development, 8*, 93-111.

Koenig, M. A., Clément, F., & Harris, P. L. (2004). Trust in testimony: children's use of true and false statements. *Psychological Science, 15*, 694-698.

Koenig, M. A., & Harris, P. L. (2005). Preschoolers mistrust ignorant and inaccurate speakers. *Child Development, 76*, 1261-1277.

Putnam, H. (1975). The meaning of "meaning." *Mind, Language, and Reality* (pp. 215-271). New York: Cambridge University Press.

Sabbagh, M. A. & Baldwin, D. A. (2001). Learning words from knowledgeable versus ignorant speakers: Links between preschoolers' theory of mind and semantic development. *Child Development, 72*, 1054-1070.

Sobel, D.M., & Corriveau, K. H. (2010). Children monitor individual's expertise for word learning. *Child Development, 81*, 669-679.

Sloutsky, V.M. (2009). Theories about "theories": where is the explanation? Comment on Waxman and Gelman. *Trends in Cognitive Sciences, 13*, 331–332.

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., Kloos, H., & Fisher, A. V. (2007a). When looks are everything: Appearance similarity versus kind information in early induction. *Psychological Science, 18*, 179-185.

Sloutsky, V. M., Kloos, H., & Fisher, A .V. (2007b). What's beyond looks? Reply to Gelman and Waxman. *Psychological Science, 18*, 556-557.

Sloutsky, V. M., Lo, Y-F., & Fisher, A. V. (2001). How much does a shared name make things similar? Linguistic labels, similarity and the development of inductive inference. *Child Development, 72*, 1695-1790.

Waxman, S.R. & Gelman, S.A. (2009). Early word-learning entails reference, not merely associations. *Trends in Cognitive Sciences. 13*, 258-263

	0	1	2	3	4
Reliable (n =16)	.25	.06	.31	.06	.31
Unreliable (n = 17)	0	0	.18	.24	.59
No label (n= 28)	.04	.11	.07	.14	.64

Table 1. The proportion of children in each condition scoring 0, 1, 2, 3 or 4

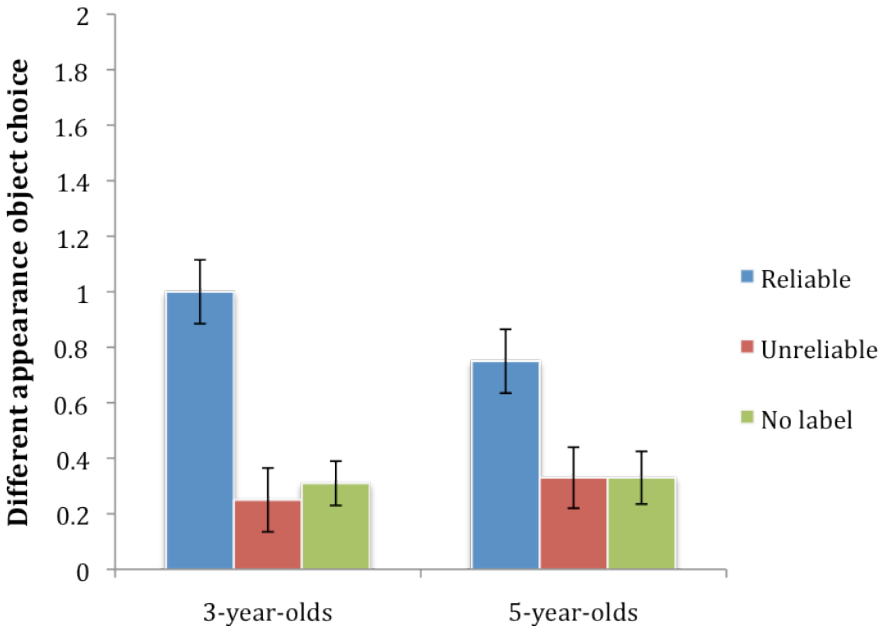


Figure 1. Children's selection of same label/different appearance objects (and the different appearance objects in the no label condition) in the first inference task.

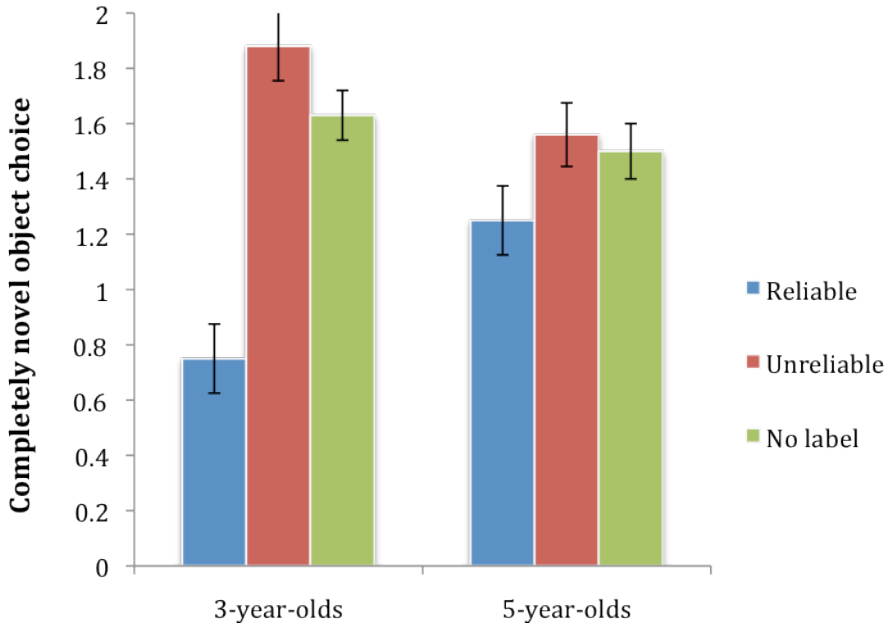
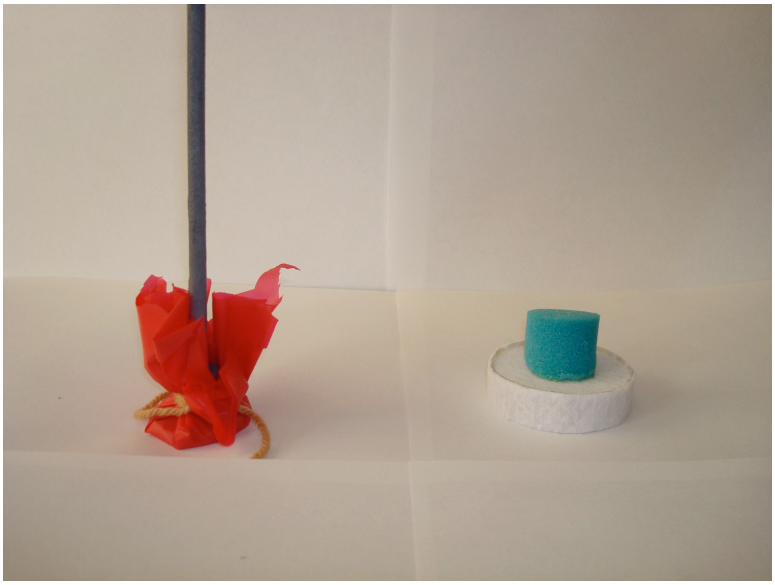


Figure 2. Children's selection of completely novel objects in the second inference task

Appendix A

An example of stimuli used in the study (the property of being magnetic)

Target (middle), same label/different appearance object (left), and same appearance/different label object (right) in the first inference task



High similar (left) and Completely novel object (right) in the second inference task