A conceptual basis for children's inductive reasoning? Evidence from lexical flexibility

Hugh Rabagliati[1]

Mahesh Srinivasan[2]

[1]University of Edinburgh

[2]University of California, Berkeley

Order of authors is arbitrary, and correspondence can be addressed to either author:

Hugh Rabagliati
School of Philosophy, Psychology and Language Sciences
University of Edinburgh
Edinburgh, UK
hugh.rabagliati@ed.ac.uk

Mahesh Srinivasan
Department of Psychology
University of California, Berkeley
Berkeley, CA
srinivasan@berkeley.edu

Data and analyses can be found at
https://osf.io/xz8uh/?view_only=b9c8fe2757e94458ac80e83c9de62a0a

Keywords: category-based induction; lexical flexibility; polysemy

Abstract

Linguistic labels clearly influence children's categorization and inductive reasoning, but it is controversial if this reflects a low-level perceptual mechanism or high-level conceptual mechanism. Two experiments (n=193 three- and four-year-olds and 49 adults) tested these accounts using the phenomenon of lexical flexibility: Words with multiple distinct but related senses (e.g., *chicken* can refer to an animal or meat). Low-level accounts predict that flexible labels may lead children – but not adults – to make seemingly "implausible" inductive inferences, like inferring that properties of "chicken" meat may apply to "chicken" animals. However, we found that under tightly-controlled conditions, shared labels did not lead participants to draw implausible inferences. This suggests that labels influence induction via a high-level mechanism. We discuss implications for theories of conceptual and lexical development.

125 words

A conceptual basis for children's inductive reasoning? Evidence from lexical flexibility

As adults, our knowledge of the world is embedded in a complex system of conceptual categories, which we communicate about using a lexicon of labels (Carey, 1985; Gelman & Markman, 1986; Keil, 1992; Markman, 1991; Murphy, 2004; Osherson, Smith, Wilkie, Lopez, & Shafir, 1990). We use labels like *chicken*, *bird*, and *mammal* to pick out specific kinds of things, and we understand that the members of these categories share important properties that go beyond their perceptual similarity. These properties of labels and categories support *inductive inference*: The ability to extend our existing knowledge to novel exemplars. For example, if we believe that chickens eat worms, we can reason that something else that has been called "a chicken" will also eat worms, even if it looks nothing like other chickens that we have encountered. Many studies in developmental science have found that children's inductive inferences are strongly affected by whether two items share a common label, e.g., whether they are each called "chicken" (Davidson & Gelman, 1990; Deng & Sloutsky, 2013; Gelman, 2003; Gelman & Coley, 1990; Gelman & Davidson, 2013; Gelman & Heyman, 1999; Gelman, Hollander, Star, & Heyman, 2000; Gelman & Markman, 1986, 1987; Gelman & Waxman, 2007; Gelman & Wellman, 1991; Heyman & Gelman, 1999; Lopez, Gelman, Gutheil, & Smith, 1992; Sloutsky & Fisher, 2011, 2012; Sloutsky, Lo, & Fisher, 2001, see Figure 1 for details). However, there has been considerable debate over what this effect of shared labels reveals about the nature of children's early conceptual representations. On the one hand, some have argued that children – like adults – use shared labels to infer that disparate items are members of a common conceptual category and are thus likely to share properties (*conceptual theories*; Gelman, 2003; Gelman & Davidson, 2013). On the other hand, others have suggested that for children, a shared label merely increases perceived similarity among items, and that it is these similarity judgments – rather than inferences based on knowledge of conceptual categories – that form the basis for children's early inductive inferences (similarity based theories; Sloutsky & Fisher, 2004; Sloutsky & Fisher, 2011, 2012).

As we will describe below, it has proven difficult to disentangle conceptual and similarity-based theories of children's inductive inference. Here, we test the predictions of these different accounts by leveraging an underexplored feature of words – the fact that many words can be used flexibly, to label items from multiple distinct but related categories (Caramazza & Grober, 1976; Klein & Murphy, 2001; Nunberg, 1979; Srinivasan & Rabagliati, 2015). Consider the word *chicken*, which has multiple different senses that can refer to an animal (*thirsty chicken*), to the meat derived from that animal (*grilled chicken*), to a scared person (*nobody calls me chicken*), and even to a toy (*rubber chicken*). As adults, we understand that while the properties of one chicken animal (e.g., that it likes to eat grain) are likely to be shared with another chicken animal, these properties are unlikely to be shared by chicken meat or rubber chickens, even though these items can all be called "chicken". The present studies explore whether children share this intuition, as predicted by conceptual theories in which children understand that chicken animals and meat are different kinds of things, or if instead shared labels make children more likely to erroneously extend properties between items from different categories (e.g., thinking that an animal called "chicken" will share properties with meat called "chicken") as a result of a change in their perceived similarity, as predicted by similarity-based theories.

FIGURE 1 ABOUT HERE

The Role of Labels in Inductive Reasoning

If a 4-year-old is shown a picture of an unusual looking bird and told that "this bird's legs get cold at night", and is then shown a picture of a bat and told that "this bat's legs stay warm at night"*,* they will generally infer that a more typical looking "bird" will also have cold legs at night, even if that bird is more perceptually similar to the "bat" than to the first, atypical "bird" (Gelman & Markman, 1986, Figure 1). In a series of papers, Susan Gelman and colleagues (Davidson & Gelman, 1990; Gelman, 1989; Gelman, 2003; Gelman & Davidson, 2013; Gelman & Markman, 1986, 1987; Heyman & Gelman, 2000; Lopez et al., 1992) have argued that this finding from the 'triad' inductive inference paradigm indicates that children and adults use similar processes to reason inductively. In particular, children and adults are thought to understand that individual labels pick out categories, and to base their inductive inferences on their knowledge of those categories. Thus, if two items are both called "bird", then they may come from the same category, and so are likely to have similar properties, like having cold legs. In contrast, if the two items receive distinct labels ("bird" vs. "bat"*),* then they may come from different categories (provided that they are not synonyms), and so may have different properties.

Consistent with this conceptual view of children's inductive inferences, children's use of labels in triad tasks appears to be 'smart', and to vary as a function of the specific properties and categories that children are asked to reason about. For example, children rely *less* on shared labels when reasoning about properties that are unlikely to be shared by category members, such as transient or externally-caused properties (e.g., "fell on the floor" rather than "has pectin inside"; Gelman, 1988). Children also rely *more* on shared labels when reasoning about items that belong to newly-learned categories that have a stronger, more natural 'conceptual basis' (e.g., reasoning about newly learned insects whose lifecourse and origins are also described; Davidson & Gelman, 1990; Gelman & Davidson, 2013), as opposed to more ad-hoc or arbitrary categories that are learned with minimal background knowledge. Together, these considerations suggest that children's inductive inferences reflect a sophisticated process of reasoning over well-structured, adult-like conceptual representations.

But this claim – that children possess adult-like conceptual knowledge, which they use to reason inductively – has proved controversial. In particular, Sloutsky, Fisher and colleagues (Deng & Sloutsky, 2012; Deng & Sloutsky, 2013; Sloutsky & Fisher, 2011, 2012; Sloutsky, Fisher, & Kloos, 2015; Sloutsky et al., 2001) have argued that young children's induction does not leverage domain-specific knowledge of concepts and categories (e.g., recognizing the importance of biological features like internal organs when reasoning about natural kinds), and have instead proposed that preschoolers represent the world based on relatively simple similarity metrics, and only build stores of domain specific knowledge later in childhood. This theory presumes that preschoolers' inductive inferences cannot be made on the basis of rich conceptual knowledge – i.e., a judgment that two things *are members of the same kind* – but instead must be based on perceived similarity (e.g., a judgment that two things are similar, Fisher & Sloutsky, 2005; Sloutsky & Fisher, 2004).

This similarity-based account explains the shared label effect through a low-level mechanism, by which linguistic labels contribute to the similarity that children perceive between items (Sloutsky & Fisher, 2012; Sloutsky et al., 2001). For example, when children compare two items that are both called "bird" (Figure 1), they may perceive them as being more similar than if they had been given distinct labels like "bird" and "bat". Because similarity is argued to provide a basis for inductive inference, children should thus be more likely to generalize properties between two items called "bird" compared to when the same items are called "bird" and "bat". Critically, modeling work by Sloutsky, Fisher, and colleagues (Sloutsky et al., 2001, Sloutsky & Fisher, 2004) suggests that labels are likely to affect inductive inference in this way even when the compared items are visually very dissimilar (e.g., as in Figure 1). For example, Sloutsky et al. (2001) estimate that having a shared label makes the perceived similarity of two items up to ten times greater than it would have been had those items received different labels. This predicts that shared labels should have powerful effects on children's responses in inductive inference tasks, as well as in other tasks that rely on perceived similarity.

Sloutsky, Fisher, and colleagues have provided a wide range of evidence in support of their position. For example, they have shown that children's similarity judgments are affected by whether items share labels (Sloutsky & Fisher, 2004; Sloutsky & Lo, 1999) and that children's inductive inferences vary as a function of similarity (Sloutsky & Fisher, 2004, 2011; Sloutsky et al., 2001). Additionally, they have provided alternative accounts of many of Gelman's findings. For instance, they have shown that while children can learn new categories that have a relatively natural "conceptual basis", they may still reason about those categories using similarity (Sloutsky, Kloos, & Fisher, 2007). They have also shown that children are more likely to generalize properties between two animals if their names are phonologically more similar (suggesting that labels contribute to similarity in a graded fashion; Sloutsky and Fisher, 2012), and have suggested that low-level linguistic statistics might affect induction (e.g., children may be more willing to generalize properties of something called a "rabbit" to something called a "bunny" because the two words are frequently collocated in child-directed speech; (Fisher, 2010; Fisher, Matlen, & Godwin, 2011). Arguing from these and other studies, these researchers maintain that true category-based induction – in which inferences go beyond the perceptual input – only emerges around 7 years of age (Badger & Shapiro, 2012; Fisher & Sloutsky, 2005; Sloutsky & Fisher, 2011; Sloutsky et al., 2015).

In sum, considerable evidence supports a similarity-based account of children's early inductive inferences, making it a serious competitor to the conceptual account. A resolution of this debate is critical for determining the nature of children's knowledge and understanding of the relation between words and concepts.

We propose that flexible uses of words could tease apart similarity-based and conceptual accounts of children's induction. In particular, under the similarity account, hearing the same flexible label applied to two items from different kinds—e.g., an animal called "chicken" and a rubber toy called "chicken"—should increase the perceived similarity between these items, and thus increase the probability that children will generalize properties between them. If this prediction is correct, it would provide important evidence against conceptual accounts of children's inductive reasoning, which hold that a shared label *should not* lead children to be more willing to generalize properties between items that they believe to be conceptually-different. According to conceptual accounts, children should only generalize properties between conceptually-different items if they have some reason to think that these properties may be shared.

To understand why the similarity-based and conceptual accounts make these predictions, consider the following propositions (square brackets indicate word sense):

(1) [Animal] *chickens* like to eat grain. Thus, [rubber] *chickens* like to eat grain.

(2) [Animal] *chickens* have skinny legs. Thus, [rubber] *chickens* have skinny legs.

For adults, the first proposition seems implausible, because we understand that rubber chickens are inanimate, and thus do not eat anything (cf. Klein & Murphy, 2002). The second proposition, in contrast, seems more plausible, because we know that rubber chickens are supposed to resemble real chickens, and thus might share properties related to appearance. Critically, similarity-based accounts like those of Sloutsky and colleagues predict that even the first inference may be plausible for young children, or at least more plausible than it is for adults. This is because 1) for children, the shared label "chicken" should increase perceived similarity between the animal and toy, and 2) children should rely more on perceived similarity than adults when drawing inductive inferences. Thus, children – but not adults – should make more implausible inductive inferences when they hear two distinct kinds of items receive a common, flexible label.

The conceptual account, in contrast, predicts that children should not be more likely to make implausible inferences when items share a common flexible label, so long as they can determine whether the label is referencing the same or different concepts across its different uses. This position is suggested by Gelman and Kalish (2006), who write (p. 696):

Is the relevant factor the label per se, or does the label work as a cue because it activates other assumptions, such as essentialism? We would argue the latter. One problem with assigning too central a role to language is that not all names promote inductive inferences. Children learn homonyms (Lily as a name versus lily as a flower), adjectives (sleepy), and nonkind nouns (pet), and these words do not seem to work in the same way as category labels such as "bird."

Thus, the conceptual account predicts that children – like adults – should not be more likely to generalize properties between conceptually-different items that share a common flexible label, provided that they can recognize whether the label is being used to refer to items from distinct kinds.

Prior research suggests that young children have a sophisticated understanding of lexical flexibility. For example, three- and four-year-olds can use the linguistic context around a flexible word to determine which of its senses are being used (e.g., whether *chicken* refers to an animal or meat; Khanna & Boland, 2010; Rabagliati, Marcus, & Pylkkänen, 2010; Rabagliati, Pylkkänen, & Marcus, 2013; Srinivasan & Snedeker, 2014). Children also understand how the senses of flexible words are conceptually related, which suggests that they may appreciate how word meanings are embedded in a broader system of world knowledge. For instance, four-year-olds can use one sense of a word to infer its other senses (e.g., if *daxing* labels an action involving a tool, then *dax* will label the tool itself, similar to how one can *shovel* snow with a *shovel*; Srinivasan, Al-Mughairy, Foushee, & Barner, 2017), and distinguish between flexible words whose distinct senses are related (e.g., words like *chicken*) and homophones that have unrelated meanings (e.g., baseball vs. animal *bat*, Srinivasan & Snedeker, 2011). In sum, children who hear a flexible word are readily able to determine which of its senses are being used and how they are related. Here, we explore whether children are able to use such knowledge when drawing inductive inferences, to recognize when the referents of flexible labels are likely to share—or not share—properties.

Our Approach

As reviewed above, the similarity-based and conceptual accounts make different predictions about how flexible uses of words might affect children's inductive reasoning. To test these predictions, we used a modified version of the triad inductive reasoning task introduced by Gelman and Markman (1986, Figure 2). As in their task, we asked child and adult participants whether a target picture was more likely to share the property of a visually similar picture, or the property of a visually dissimilar picture. We varied whether or not the dissimilar picture was named by the same label as the target (e.g., both were called "chicken") and, critically, also varied the conceptual relation between the dissimilar picture and target. In the "Same Sense" condition, the dissimilar and target pictures were drawn from the same sense of a flexible label (e.g., two different looking animals called "chicken"). In the "Different Sense" condition, the dissimilar and target pictures were drawn from two different senses of the flexible label (e.g., an animal called "chicken" and meat called "chicken"). Our studies measured participants' willingness to base their inductive inferences on a shared label, and whether this depended on the sense(s) in which the label was being used (Figure 2).

FIGURE 2 ABOUT HERE

If the effect of labels on children's inductive inferences is *conceptually-driven*, then the application of a shared label to the dissimilar and target pictures may affect inferences in the Same Sense version of this task, because this shared word sense could allow children to override perceptual similarity and reason based on shared category (e.g., by recognizing that the perceptually-different animals are both chickens and likely to share properties). However, the application of a shared label to the dissimilar and target pictures should not affect children's inferences in the Different Sense version of the task (e.g., between an animal called "chicken" and meat called "chicken"), so long as children can recognize that the label is being used with different senses and thus does not implicate a shared category. In contrast, if the effect of shared labels on children's inductive inferences is based on *similarity*, then children should show an effect of shared labels not just in the Same Sense version of the task, but also in the Different Sense version of the task, because the shared label should increase perceived similarity between the dissimilar and test pictures in both conditions, regardless of the sense(s) in which it is used. Finally, both the conceptual and similarity theories predict that adults should only show an effect of shared label in the Same Sense condition, because both theories assume that adults use conceptual categories—and do not rely on shared labels alone—to make inductive inferences (Badger & Shapiro, 2012; Fisher & Sloutsky, 2005; Sloutsky & Fisher, 2011; Sloutsky et al., 2015; Sloutsky & Lo, 1999).

Experiment 1 explored how the inductive inferences of children and adults varied between the Same Sense and Different Sense versions of this task, and how the presence of a shared label affected these inferences. To test the effect of shared labels, we varied within subjects whether pictures were given category-marking labels (e.g., such that the dissimilar and target pictures were each called "chicken" and the similar test was called "duck".), or were simply referred to with the demonstrative pronoun *this*.

Experiment 1

Methods

Participants

Our final sample included 97 children, 48 3-year-olds (*M*=43 months; 36-47 range; 22 female) and 49 4-year-olds (*M*=54 months; 48-59 range; 23 female). Prior studies on the triad task have used approximately 20 children per condition; we more than doubled this number as we were aiming to test for an interaction effect. Forty-nine children were run in the Same Sense task, and 48 in the Different Sense task. Fifty-five children were tested in Edinburgh, UK, at local preschools and in lab, and 42 children were tested in the Berkeley, California area, at local museums, preschools and in lab. Eight additional children were tested but excluded for failing to complete the task (3), having developmental delays (2), or due to experimenter error (3). An additional 24 adults were also tested (*M*=22 years; 18-31 range; 14 female), 10 at the University of Edinburgh and 14 at UC Berkeley. All participants spoke English as their primary language. We did not record ethnicity or social background. Informally, the Edinburgh sample was typically White and middle class, while the Berkeley sample was drawn from a variety of ethnicities and was typically middle class.

Materials

We created twelve triads based on three different flexible words (see Figure 2). *Chicken* triads always used a chicken animal as Target, a duck as the Visually Similar Test Picture, and either a chicken animal (Same Sense task) or chicken meat (Different Sense task) as the Dissimilar Test Picture. *Glass* triads used a piece of glass material as Target, a piece of plastic as the Similar Test Picture, and either a piece of glass material (Same Sense task) or a drinking glass (Different Sense task) as the Dissimilar Test Picture. *Horse* triads used a horse animal as Target, a zebra as the Similar Test Picture, and either a horse animal (Same Sense task) or a toy horse (Different Sense task) as the Dissimilar Test Picture. For each flexible word, we created four different triads, each of which used different pictures (e.g., using different exemplars of chickens) and properties.

Properties were chosen with the intuition that adult participants would prefer to generalize from the Target picture to the Dissimilar Test picture in the Same Sense condition (e.g., from one chicken animal to another) but would prefer not to do so in the Different Sense condition (e.g., from chicken meat to a chicken animal). However, we also made certain that, in both the Same Sense and Different Sense tasks, the properties could in principle apply to all of the different pictures (e.g., to a chicken animal, chicken meat, and a duck). This meant that we could not use certain deep, conceptual properties that differentially apply to animate vs. inanimate items (e.g., *has dax in its blood*). Our final set of properties (see Supplemental Material) described attributes that could not be perceived from the pictures themselves, such as *people keep this chicken warm and they keep this duck cold,* or *this horse has some dax in it and this zebra has lots of wug in it*. These descriptions mostly used familiar terms, but occasionally used novel terms like *dax*.

Procedure

To ensure that participants were making inference judgments under some uncertainty, we framed the task as a game in which participants had to guess answers to questions about objects found on a planet called Jupp.

On each trial, the triad of pictures was displayed to the participant on a laptop computer screen, and the experimenter read the properties aloud, following a consistent script. Below is an example script of how pictures were labeled:

Do you see this chicken [point to Target]? And do you see this duck [point to Similar Test picture]? And do you see this chicken [point to Dissimilar Test picture]? On planet Jupp, people like to feed this duck [point to Similar Test], and they like to sell this chicken [point to Dissimilar Test]. What about this chicken [point to Target]? Do they like to feed it, like this duck [point to Dissimilar Test], or do they like to sell it, like this chicken [point to Similar Test]?

Participants then judged whether the Target had the property of the Dissimilar or the Similar Test picture, by either naming or pointing at the picture.

Between subjects, we varied whether participants took part in the Same Sense or Different Sense conditions. Within subjects, we varied whether participants heard category labels for each picture of a triad (e.g. *do you see this chicken?*), or just demonstrative pronouns (*do you see this?*). We counterbalanced order of mention of the Dissimilar and Similar Test pictures, as well as the left-right positions of these pictures on the screen. Participants received trials in one of two random orders.

Participants always completed two simple Same Sense trials as a warm up, before the 12 test trials (see Supplemental Material).

Analysis

We analyzed the proportion of trials in which participants chose to extend the property from the Dissimilar Test Picture to the Target, e.g., for the Chicken triads, whether participants chose the chicken animal (Same Sense task), or the chicken meat (Different Sense task), as opposed to the duck. We used mixed effects logistic regressions to separately model the choices of children and adults. Our dependent variable was participants' choice (Dissimilar Test or Similar Test Picture) and our predictor variables were Induction Task (Same or Different Sense), Label Presence (Shared Label or No Label), and their interaction. We included the maximal random effects structure that permitted convergence. In practice, this was random intercepts for each subject, a by-subject random slope for Label Presence, and random intercepts for each item (chicken, glass, and horse). Preliminary analyses did not find significant effects of age for children, or of testing location (Edinburgh vs. Berkeley), and so these factors were dropped from subsequent models.

Results

On the one hand, if children's inductive inferences are mediated by intuitions about category membership, then children and adults may show effects of shared label in the Same Sense version of the task, where the shared label (unlike the demonstrative pronoun) may help indicate that the Dissimilar picture (e.g., a canonical chicken animal) is from the same taxonomic category as the Target (an unusual-looking chicken animal; Figure 2). However, neither children nor adults should show effects of shared label in the Different Sense version of the task, because they can recognize that the Dissimilar Test picture (e.g., chicken meat) is a member of a different category than the Target (a chicken animal), even if the two items are given a common, flexible label (they are both called "chicken"). On the other hand, if children – unlike adults – make inductive inferences on the basis of similarity, and if shared labels act to increase perceived similarity, then the presence of a shared flexible label should affect children's but not adults' inductive generalizations even in the Different Sense version of the task (see Supplemental Material for full descriptive statistics and tests against chance).

Consistent with the similarity account, children were more likely to choose the Dissimilar Test picture when it shared a label with the Target (Figure 3, Beta=-0.68 (Standard Error=0.14), Wald's Z=4.94, p<.001), and this effect did not vary between the Same Sense and Different Sense versions of the task (i.e., there was no Label Presence by Induction Task interaction: Beta=0.03(0.19), Z=0.17, p=.86). Follow-up analyses showed that Label had a significant effect in both the Same Sense version of the task (Beta=-0.70(0.21), Z=3.3, p<.001) and the Different Sense version of the task (Beta=-0.66(0.18), Z=3.7, p<.001), and visual inspection of the results for the three types of triads (Chicken, Glass, Horse) suggests that the effect of Label was consistent across items (see Figure 4). In addition, children were also more likely to choose the Dissimilar Test pictures in the Same Sense version of the task than in the Different Sense version (Beta=0.40(0.12), z=3.5, p<.001), a result that is predicted under all accounts, since the Dissimilar Test pictures in the Different Sense task were less perceptually and conceptually similar to the Target pictures than they were in the Same Sense task.

Together, these findings from children are consistent with the idea that early inductive inferences are based on similarity, rather than on conceptual reasoning. In particular, shared labels may have increased perceived similarity between the Dissimilar Test and Target items in both conditions, leading children to make implausible inferences in the Different Sense task, and generalize properties between the referents of different word senses (e.g., between meat called "chicken" and an animal called "chicken").

However, data from adults showed a surprising pattern that cast doubt on this initial conclusion. In particular, adults, like children, were *also* more likely to choose the Dissimilar

Test picture when it shared a label with the Target (Beta=-1.5(0.4), z=3.8, p<.001), and again this effect was not qualified by an interaction between Label and Task (Beta=0.39(0.52), z=0.75, p=.45; Figure 3). Like children, adults were also less likely to choose the Dissimilar Test pictures in the Different Sense version of the task than in the Same Sense version (Beta=-1.3 (0.3), z=4.0, p<.001). Figure 4 shows that the effect of Label was present in all items, but was largest for the *glass* trials.

Critically, the fact that labels affected adults' performance in the Different Sense task is unexpected for *both* the similarity-based and conceptual accounts of children's inductive inference. This is because both theories assume that adults use labels to make inferences about category membership, but ultimately rely on their knowledge of categories to make inductive inferences (Badger & Shapiro, 2012; Fisher & Sloutsky, 2005; Sloutsky & Fisher, 2011; Sloutsky et al., 2015; Sloutsky & Lo, 1999). Thus, on both theories, adults should not have been more likely to generalize properties between items that they believed to be of distinct categories when those items received a common label, as we found in the Different Sense task. This surprising finding suggests that the effect of shared label observed in children may not reveal that children's inductive inferences are based on similarity; Instead, some other factor may have caused children and adults alike to show effects of shared label across the Same Sense and Different Sense tasks.

FIGURE 3 ABOUT HERE

Discussion

Experiment 1 suggests that children use the presence of a shared label to guide their inductive generalizations, even in contexts in which such generalizations might seem implausible, e.g., when considering whether properties of meat called "chicken" apply to an animal called “chicken”. On first glance, this result appears inconsistent with conceptual accounts of children’s inductive inferences, and supports similarity-based accounts. However, Experiment 1 also showed that children’s “implausible” inferences were, in fact, somewhat adult-like: Adults, like children, were also more likely to extend properties between conceptually-different items in the Different Sense task, when those items received a common, flexible label.

Why were adults more likely to generalize properties between items when they received shared labels in the Different Sense task, and what might this suggest about the nature of children’s inductive reasoning? One possibility is that for both adults and children, the application of a shared label increases perceived similarity, providing a basis for inductive inferences. However, this conclusion is inconsistent with a previous consensus in the literature that adults’ inductive reasoning is not similarity-based, but is instead conceptually driven (Badger & Shapiro, 2012; Fisher & Sloutsky, 2005; Osherson et al., 1990). This raises the possibility that some other factor, particular to the design of Experiment 1, may have caused both children and adults to rely on shared labels in the Different Sense condition.

From the participants’ point of view, one perhaps unusual feature of Experiment 1 was that we used a within-subjects manipulation of label: The experimenter only provided labels for the pictures on half of the trials (e.g., using “chicken” to label both the Dissimilar Picture and the Target, and “duck” to label the Similar Picture) but used pronouns on the other half of trials (e.g., using “this” to label the pictures). The simultaneous use of these two referential strategies may have led participants to draw a pragmatic inference about the experimenter’s intentions (McGarrigle & Donaldson, 1974): *Perhaps the experimenter wants me to attend to the shared labels on this trial, since she does not use labels on all trials, and this must be for a reason.* Critically, participants could have drawn this pragmatic inference in both the Same Sense *and* Different Sense versions of the task, leading to the effect of shared label we observed in both tasks, in both children and in adults. Experiment 2 tested if the effect of shared label would still hold when this inference could no longer be drawn.

Experiment 2

To limit the potential pragmatic effects of exposing participants to two referential strategies, in which labels are used on some trials but not others, we modified our experimental design to use labels on all trials. In particular, instead of contrasting shared labels with demonstrative pronouns, we contrasted shared labels with synonyms (e.g., using “chicken” to label both the Dissimilar picture and the target on some trials, and using “hen” or “drumsticks” to label the Dissimilar picture and “chicken” to label the Target on other trials; see Figure 2). This manipulation was guided by a result from Gelman and Markman (1986) who showed that children can use synonyms as a basis for inductive inferences. They found that, just as children transfer properties between two dissimilar items called “rabbits” (rather than between a perceptually-similar “rabbit” and “squirrel”), children also transfer properties between dissimilar items called “rabbit” and “bunny”, suggesting that children’s inferences are not made on the basis of shared labels, *per se*, but instead on the basis of their meanings (but see Fisher, 2010 and Fisher et al., 2011, who have offered a different interpretation of these findings that we return to in the General Discussion).

If the effect of shared labels in Experiment 1 was caused by a pragmatic inference about the experimenter’s use of labels on some trials but not others, then this effect should not arise in Experiment 2, when labels are used on all trials. But if the effect of shared labels in Experiment 1 reflects similarity-based reasoning – in which shared labels increase perceived similarity and thus support inductive inference – then it should still be present in Experiment 2, in both the Same Sense and Different Sense conditions, since synonymous labels like *hen* and *chicken* are perceptually distinct, and thus should *not* increase the perceived similarity of the Dissimilar picture and the Target picture for children.

Methods

Participants

Our final sample included 96 children, including 48 3-year-olds (*M*=42 months; 36-47 range; 27 female) and 48 4-year-olds (*M*=51 months; 48-59 range; 22 female). Half of the children in each age group received the Same Sense task, and the other half received the Different Sense task. Forty-five children were tested in the Edinburgh, UK area, at local preschools and in lab, and 51 children were tested in the Berkeley, California area, at local museums, preschools and in lab. Three additional children were excluded, due to inattention during the task (1), parental interference (1), or because they did not complete the task (1). An additional 25 adults were also tested, all at the University of Edinburgh. All participants had English as a primary language.

Materials, Procedure and Analysis

We used the same materials and procedure as in Experiment 1, except that we replaced the pronoun trials with trials in which all pictures were labeled, such that the Target and Dissimilar Test picture were labeled with synonyms. As before, on Shared Label trials, the Dissimilar Test picture and the Target were given the same flexible label (e.g., “chicken”, “glass”, etc.). But on Different Label trials, the Dissimilar Test picture was labeled with a synonym of the label given to the Target (e.g., “hen” for the Same Sense version of the chicken triad, and “drumsticks” for the Different Sense version; see Supplemental Material).

We followed a similar analytic strategy to Experiment 1. The dependent variable in our mixed effects logistic regressions was participants’ choice (Dissimilar or Similar Test pictures) and our predictor variables were Induction Task (Same or Different Sense), Label Type (Shared Label or Synonym), and their interaction. We included the maximal random effects structure that permitted convergence: random intercepts for each subject, a by-subject random slope for Label Type, random intercepts for each item (chicken, glass, and horse) and a by-item slope for Task.

Results

As in Experiment 1, inductive reasoning was affected by the task that children and adults received (Figure 4; see Supplemental Material for full descriptive statistics): Participants were overall more likely to extend the Dissimilar Test picture property in the Same Sense version of the task (e.g., between two chicken animals) than in the Different Sense version of the task (e.g., between chicken meat and animal, Children: Beta=-0.49(0.16), Z=3.0, p=.002, Adults: Beta=-3.57(0.79), Z=4.49, p<.001), presumably because the Dissimilar and Target items were more dissimilar and less related in the Different Sense task than they were in the Same Sense task.

Strikingly, however, with the inclusion of synonyms in Experiment 2, we no longer observed any reliable effect of shared label. This was true both for children (Main effect of Label: Beta=-0.12(0.09), Z=1.35, p=.18; Interaction of Label and Task: Beta=0.21(0.13), Z=1.6, p=.10), and also for adults (Main effect of Label: Beta=-0.80(0.53), Z=1.51, p=.13; Interaction of Label and Task: Beta=0.93(0.79), Z=1.17, p=.24). Visual inspection of the results for the three different types of triad (chicken, glass, horse; Figure 4) suggests that the effects found in Experiment 2 were consistent across items.

Thus, unlike in Experiment 1, shared labels did not lead our child or adult participants to make more implausible inferences in the Different Sense task, e.g., in which properties of meat called “chicken” were generalized to an animal called “chicken”. A Bayesian analysis confirmed this. Using the BayesFactor package (Rouder & Morey, 2015), we compared whether the effect of shared labels for the Different Sense task was better explained by a null hypothesis (no effect of label) or an alternative hypothesis in which label has an effect; A Bayes Factor less than $\frac{1}{3}$ is evidence for the null, and over 3 is evidence for the alternative. For both groups, the evidence favored the null (Adults: B.F. = 0.28, Children: B.F. = 0.16).

Finally, a post-hoc test revealed that, for children in the Same Sense task, shared labels caused a reliable increase in the proportion of dissimilar test choices (Beta=-0.27(0.13), Z=-2.0, p=.04). One possible reason for this is that – in contrast to adults – children may have struggled to use synonyms in the Same Sense task (e.g., *chicken* and *hen*) to infer that the Target and Dissimilar Test picture were from the same category. This could be because many children may not have known the synonyms we used, as some of them were low in frequency (see Supplemental Material). By contrast, adults would have known these synonyms, perhaps explaining why they did not show an effect of shared label (Beta=-0.6(0.51), Z=-1.2, p=.24).

FIGURE 4 ABOUT HERE

Comparison of Experiments 1 and 2

When the pragmatic demands to use shared labels were reduced in Experiment 2, shared labels no longer made adults or children more likely to make implausible inferences, and generalize properties between the referents of different word senses. To confirm this difference, we compared the two experiments statistically. We used mixed effects logistic regressions with Choice (Dissimilar or Similar Test picture) as the dependent variable, and Task (Same or Different Sense), Label Type (Shared or Not shared), Experiment (1 or 2) and their full set of interactions. We included the maximal random effects structure that permitted convergence: Random intercepts for subjects and by-subject slopes for label, and random intercepts for each item (chicken, glass, and horse) and by-item slopes for task.

The results for both children and adults were very similar. Both groups showed effects of Task, choosing the Dissimilar Test picture more in the Same Sense version of the task (Children: Beta=-0.33(0.09), Z=3.74, p<.001; Adults: Beta=-1.13(0.16), Z=6.99, p<.001), and they also both showed effects of Label, choosing the Dissimilar Test picture more often when it shared a label with the Target (Children: Beta=-0.30(0.07), Z=4.51, p<.001; Adults: Beta=-0.60(0.12), Z=5.1, p<.001; Adults were also less likely to choose the Dissimilar test picture in Experiment 1, i.e., a main effect of Experiment, Beta=-0.42(0.16), Z=2.7, p=.008). But critically, the effect of Label in both age groups was qualified by an interaction with Experiment: Labels had a robustly reduced effect in Experiment 2 compared to in Experiment 1, for both children (Beta=-0.18(0.07), Z=2.77, p=.004) and for adults (Beta=-0.50(0.12), Z=4.24, p<.001).

Discussion

While Experiment 1 suggested that children and adults make more implausible inferences across different senses of flexible words when they receive a shared label – such as generalizing properties of meat called "chicken" to an animal called "chicken" – Experiment 2 showed that this effect disappeared once the pragmatic demands of the task were changed, such that labels were used on all trials. Experiment 2 therefore suggests that the effect of shared labels in Experiment 1 was not caused by children's use of similarity for inductive inferences, but instead may have resulted from a belief that the experimenter wanted them to make use of the shared label information that was provided on only a select subset of trials.

General Discussion

What is the relation between words and concepts during development? Do children understand that words label categories that are situated within a well-structured system of world knowledge? Or are words more akin to perceptual features, associated with an unstructured assortment of other features, in which categories are only defined by their similarity? Here, we leveraged the phenomenon of lexical flexibility – words with multiple related senses, such as *chicken* – to tease these two accounts apart.

Similarity-based accounts predict that shared labels should affect how young children make inductive inferences, even when that label is used to refer to two different kinds of things, e.g., such that they are more likely to generalize the properties of meat to an animal when they have each been called "chicken". The key results of our studies did not support this prediction. While Experiment 1 found that children are more likely to draw implausible inferences between conceptually-different items that receive a common label, it also showed that adults do the same, which suggests that the inferences made by both groups were probably the result of a pragmatic bias: Participants may have assumed that the experimenter wanted them to use the shared label information, which was presented on some trials but not others. Consistent with this, when the pragmatic demands to use shared labels were reduced in Experiment 2, neither children nor adults were more likely to draw implausible inferences in the Different Sense condition when shared labels were provided. From these data we conclude that, by the preschool years, children can use their understanding of the complex relations between word senses and concepts to guide their inductive inferences, and thus recognize when two things that are called "chicken" are likely to share the same properties, and when they are not. Below, we discuss the implications of our findings for both similarity-based and conceptual accounts of induction, as well as for theories of lexical development.

Implications for similarity-based accounts of inductive inference

We believe that our results and conclusions are robust to a variety of critiques that proponents of a similarity-based view might advance. For example, it could be argued that, given our experimental design, similarity-based theories would predict that labels should *not* have affected children's inferences in the Different Sense task of Experiment 2. In particular, children in that task could have perceived the Dissimilar Test pictures (e.g., a picture of chicken meat) to be *so different* from the Target pictures (e.g., an unusual looking chicken), that shared labels could not have increased perceived similarity enough to affect children's inductive inferences in a measurable way.

One reason to doubt this alternative account comes from prior implementations of Sloutsky and Fisher's (2004) SINC model, which suggest that the effect of shared label on children's judgments should be large even when items are quite perceptually dissimilar. In particular, the SINC model predicts that, even when the similarity of the Dissimilar Test picture to the Target is *78 times less* than the similarity of the Similar Test picture to the Target, the probability of choosing the Dissimilar Test picture should still increase by 10 points when it shares a label with the Target.

Applied to our data, the SINC model itself suggests that children in the Different Sense tasks did not perceive the Dissimilar Test pictures as being 78 times less similar to the Target pictures than the Similar test pictures were. If there had been a 78-fold difference, the SINC model predicts that children should only have chosen the Dissimilar Test pictures on 1% of trials when they did not share a label with the Target. However, the lowest proportion of choices of the Dissimilar Test pictures that we observed from children was 27% (Experiment 1, Different Sense task, pronoun condition). Working backwards from this 27% figure, the SINC model predicts that in the Different Sense tasks, children perceived the Similar Test pictures as being only 2.7 times more similar to the Target pictures than the Dissimilar Test pictures were. Given this, the SINC model predicts that a shared label should have increased the probability of choosing the Dissimilar Test picture in the Different Sense task from 0.27 to 0.79. Clearly, we did not observe such an increase.

Modified versions of the SINC model are also unlikely to explain our data. For example, Fisher (Fisher, 2010; Fisher et al., 2011; Godwin & Fisher, 2015) has shown that children's similarity judgments are affected not only by hearing a shared label, but also by hearing two distinct labels that frequently co-occur. For example, children often hear *bunny* and *rabbit* collocated, which could lead them to associate these two distinct words as a single feature, allowing them to function similarly to shared labels. If distinct words can function similarly to shared labels, then this could potentially account for the null effect of shared label observed in Experiment 2's Different Sense task. In that task, children heard a shared label for the Target and Dissimilar Test ("chicken" and "chicken") on some trials, and heard distinct, synonymous labels for these pictures on other trials ("chicken" and "drumsticks"). Might children treat these synonym pairs as a single feature because they have frequently heard them collocated?

To test this idea, we calculated co-occurrence statistics for the synonyms pairs used in Experiment 2 based on the British English corpora in CHILDES (MacWhinney, 2000); see Supplemental Material for methodological details). The resulting statistics suggest that it is unlikely that collocations of the synonyms have led them to be treated as a single feature, particularly in the Different Sense task. Association levels between the synonyms were extremely low overall and were in fact lower in the Different Sense than in the Same Sense task (Different Sense: $M = 0.0016$; Same Sense: $M = 0.0025$). Overall, they were an order of magnitude lower than the association levels between Fisher (2010)'s synonyms (M = 0.03, but see Supplemental Material for differences in our methods).

Together, these considerations make it unlikely that a simple similarity-based mechanism could account for how children generated inductive inferences in Experiment 2.

Following others (Davidson & Gelman, 1990; Gelman & Davidson, 2013; Gelman & Waxman, 2007), we suggest that similarity-based accounts of children's reasoning should acknowledge an early-developing capacity for representing at least some limited knowledge of the world, as well as for understanding the referential functions of labels.

Implications for conceptual accounts of inductive inference, and for theories of semantic and lexical development

In principle, our findings provide support for conceptual accounts of inductive inference, because they show that children's inferences are based on the meanings of words rather than on the mere application of shared labels. However, our data also suggest some important ways in which current conceptual accounts may need to be modified, and provide a cautionary lesson about how to interpret effects of shared labels.

We turn to the cautionary lesson first. Our two experiments suggested that shared labels can have quite different effects on how children and adults draw inferences, depending upon the contexts in which they are used. We argued that children and adults in Experiment 1 showed a robust effect of shared labels only because of a pragmatic bias: By this account, participants reasoned that, when full labels (rather than pronouns) were provided on a select proportion of trials, then their choices on those trials should make use of the labels. Consistent with this, the effect of shared labels largely disappeared in Experiment 2, when we reduced pragmatic task demands by ensuring that labels were provided on all trials.

Could these task demands also explain other documented effects of shared labels in the literature? This would have far-reaching implications if true, given that many studies have used the "triad" inductive inference paradigm that we employed here. However, while we believe that the pragmatic bias documented in our studies is important to attend to, there is good reason to believe that its wider impact has not been large. In particular, unlike most other work in this field, our experiments manipulated label as a within-subjects factor. This is important, because the pragmatic task demands we identified only arise when participants can contrast trials in which labels *are used* with trials in which labels *are not used*. While this was the case in Experiment 1, most prior work has either manipulated shared labeling as a between-subjects factor (as in, e.g., Sloutsky and Fisher, 2004), or has not manipulated the presence of shared labels at all (e.g., Gelman and Markman, 1986).

How might our findings be incorporated into conceptual accounts of children's inductive reasoning and development more generally? On the one hand, our main demonstration – that children rely on *shared meanings* for inductive inference, rather than shared labels – is consistent with previous proposals that children's inferences are guided by conceptual categories rather than a blind focus on labels (see e.g., the quote from Gelman and Kalish, 2006, in the Introduction). On the other hand, many proponents of conceptual accounts have argued that children and even infants treat labels themselves as "placeholders" for conceptual categories, and operate under the assumption that all things that share a label should be grouped together (Fulkerson & Waxman, 2007; Gelman, 2003; Gelman & Brandone, 2010; Waxman & Lidz, 2006; Waxman & Markow, 1995; Xu, 2002); As we argue below, a strong placeholder assumption would predict – contrary to our findings – that children should fail to recognize that the referents of different uses of flexible words are members of distinct conceptual categories.

Although a placeholder assumption is thought to help children learn words and to form abstract categories that go beyond perceptual similarity, it could lead children astray when learning flexible words. In particular, if children are guided by a placeholder assumption, and assume that items that are labeled by the same flexible word will be members of a common category, they may mistakenly construct single, overly-broad categories that reflect all of the different ways in which a flexible word is used. This would predict, for example, that children would treat *chicken* animals, *chicken* meat, and rubber *chickens* as members of a common category that share properties. However, our findings show that shared labels do not lure children into making implausible inferences between conceptually-different items like "chicken" meat and "chicken" animals.

Our data thus point to an alternative approach to how children expect words to relate to concepts, in which children do not expect words to label a single category of meaning, but to instead be used flexibly, and to label items from multiple distinct but related categories (Dautriche, Chemla, & Christophe, 2016; Dautriche, Fibla, & Christophe, 2015; Srinivasan, 2016; Srinivasan & Rabagliati, 2015). The motivation for this approach is in part functional: It may be easier for children to learn a lexicon in which words are flexible – i.e., label multiple concepts in predictable ways – compared to an unambiguous lexicon in which each concept must be expressed by its own, separately learned label.

Support for this idea comes from recent findings that young children have a sophisticated understanding of how the senses of flexible words are related. When taught a novel word that labels one sense of a familiar flexible word (e.g., that *blicket* labels a physical book), preschoolers believe that novel word can also label the other sense of the word (e.g., *blicket* labels the book's intellectual content; Srinivasan & Snedeker, 2011, 2014). Four-year-olds also understand generalizations about lexical flexibility and can use these generalizations to anticipate how new words will be used. For instance, having learned that a new word labels an action involving a tool, children expect that same word to also label the tool itself, similar to how one can *shovel* snow with a *shovel* (Srinivasan et al., 2017). The present studies build on these findings to show that children not only represent how the senses of flexible words are related, but also represent how those different word senses are *distinct,* just as adults do (Klein & Murphy, 2001, 2002; Rabagliati & Snedeker, 2013). This implies that the basic architecture of the lexicon – in which labels are associated with multiple distinct, related senses – may be present from early in acquisition.

How can this claim – that children expect words to have multiple distinct but related meanings — be reconciled with the considerable extant evidence that children have a placeholder bias, and expect words to label items from a common category? We suspect that children are able to attend to shared labels in a "smart" way, such that they can distinguish between situations in which a label is being used to refer to items from the same kind (e.g., multiple "chicken" animals), as opposed to items from different kinds (e.g., some "chicken" animals and some "chicken" meat). For instance, children could separate distinct senses of flexible words by noting that they appear in different linguistic contexts (Dautriche et al., 2015; Rabagliati et al., 2013) or denote ontologically distinct kinds of things (Dautriche et al., 2016; Rabagliati et al., 2010; Srinivasan & Snedeker, 2014).

Interestingly, a bias to assume that words are flexible might also help children learn about concepts and categories, in a similar – but importantly different – fashion to the placeholder bias described above. While the placeholder bias is proposed to assist children in grouping dissimilar objects into a common category, a bias to assume that words are flexible might additionally allow children to construct relations between distinct categories. For example, as children become aware that the same words can label tools and their uses (e.g., a *hammer* vs. *to hammer*; *a brush* vs. *to brush*), or materials and artifacts made from their materials (e.g., a *tin* vs. *some tin*; *a glass* vs. *some glass*), they could make inferences about the structure of new categories. For example, children might expect that an object called a *wipe* is likely to belong to a category that is defined in part by that function (i.e., *wipes* are for *wiping*), or that an object called a *copper* belongs to a category defined in part by that material (i.e., *coppers* are made of *copper*). Thus, children's understanding of lexical flexibility may shape how they construe new categories. This is especially noteworthy given the widespread prevalence of lexical flexibility in natural language (Srinivasan & Rabagliati, 2015), which highlights the fact that in lexical development, children are often learning new word senses, as opposed to entirely new words.

Conclusions

Our findings suggest that when children use labels for inductive inference, they do so in a 'smart' fashion, that accounts for whether a word has multiple, distinct senses. These findings are not consistent with accounts of conceptual development in which children's early inferences are based on judgments about similarity or with accounts in which children assume that items denoted by a single label are members of a common conceptual category. Instead, our findings suggest that children reason using sophisticated and well-structured world knowledge, and that this system of world knowledge is linked to a lexicon that, from an early age, can represent flexible words as having multiple related and distinct meanings.

References

Badger, J. R., & Shapiro, L. R. (2012). Evidence of a transition from perceptual to category induction in 3-to 9-year-old children. *Journal of Experimental Child Psychology, 113*(1), 131-146.

Caramazza, A., & Grober, E. (1976). Polysemy and the structure of the subjective lexicon. In C. Rameh (Ed.), *Georgetown University Roundtable on Languages and Linguistics. Semantics: Theory and application* (pp. 181–206). Wasington, D.C.: Georgetown University Press.

Carey, S. (1985). *Conceptual change in childhood*. Cambridge, MA: MIT Press.

Dautriche, I., Chemla, E., & Christophe, A. (2016). Word Learning: Homophony and the Distribution of Learning Exemplars. *Language Learning and Development, 12*(3), 231-251.

Dautriche, I., Fibla, L., & Christophe, A. (2015). *Learning homophones: syntactic and semantic context matters.* Paper presented at the 40th Boston University Conference on Language Acquisition, Boston, MA.

Davidson, N. S., & Gelman, S. A. (1990). Inductions from Novel Categories - the Role of Language and Conceptual Structure. *Cognitive Development, 5*(2), 151-176.

Deng, W., & Sloutsky, V. M. (2012). Carrot Eaters or Moving Heads: Inductive Inference Is Better Supported by Salient Features Than by Category Labels. *Psychological Science, 23*(2), 178-186. doi:10.1177/0956797611429133

Deng, W., & Sloutsky, V. M. (2013). The role of linguistic labels in inductive generalization. *Journal of Experimental Child Psychology, 114*(3), 432-455.

Fisher, A. V. (2010). What's in the name? Or how rocks and stones are different from bunnies and rabbits. *Journal of Experimental Child Psychology, 105*(3), 198-212.

Fisher, A. V., Matlen, B. J., & Godwin, K. E. (2011). Semantic similarity of labels and inductive generalization: Taking a second look. *Cognition, 118*(3), 432-438.

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*(3), 583-597.

Fulkerson, A. L., & Waxman, S. R. (2007). Words (but not tones) facilitate object categorization: Evidence from 6-and 12-month-olds. *Cognition, 105*(1), 218-228.

Gelman, S. A. (1988). The Development of Induction within Natural Kind and Artifact Categories. *Cognitive psychology, 20*(1), 65-95.

Gelman, S. A. (1989). Childrens Use of Categories to Guide Biological Inferences. *Human Development, 32*(2), 65-71.

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

Gelman, S. A., & Brandone, A. C. (2010). Fast-mapping placeholders: Using words to talk about kinds. *Language Learning and Development, 6*(3), 223-240.

Gelman, S. A., & Coley, J. D. (1990). The Importance of Knowing a Dodo Is a Bird - Categories and Inferences in 2-Year-Old Children. *Developmental psychology, 26*(5), 796-804.

Gelman, S. A., & Davidson, N. S. (2013). Conceptual influences on category-based induction. *Cognitive psychology, 66*(3), 327-353.

Gelman, S. A., & Heyman, G. D. (1999). Carrot-eaters and creature-believers: The effects of lexicalization on children's inferences about social categories. *Psychological Science, 10*(6), 489-493.

Gelman, S. A., Hollander, M., Star, J., & Heyman, G. D. (2000). The role of language in the construction of kinds. *Psychology of Learning and Motivation: Advances in Research and Theory, Vol 39, 39*, 201-263.

Gelman, S. A., & Kalish, C. W. (2006). Conceptual development. In D. Kuhn & R. Siegler (Eds.), *Handbook of Child Psychology Volume 2. Cognition, Perception and Language* (pp. 687-733). New York: Wiley.

Gelman, S. A., & Markman, E. M. (1986). Categories and Induction in Young-Children. *Cognition, 23*(3), 183-209.

Gelman, S. A., & Markman, E. M. (1987). Young Childrens Inductions from Natural Kinds - the Role of Categories and Appearances. *Child development, 58*(6), 1532-1541.

Gelman, S. A., & Waxman, S. R. (2007). Looking beyond looks: Comments on Sloutsky, Kloos, and Fisher (2007). *Psychological Science, 18*(6), 554-555.

Gelman, S. A., & Wellman, H. M. (1991). Insides and Essences - Early Understandings of the Non-Obvious. *Cognition, 38*(3), 213-244.

Godwin, K. E., & Fisher, A. V. (2015). Inductive generalization with familiar categories: developmental changes in children's reliance on perceptual similarity and kind information. *Frontiers in Psychology, 6*.

Heyman, G. D., & Gelman, S. A. (1999). The use of trait labels in making psychological inferences. *Child development, 70*(3), 604-619.

Heyman, G. D., & Gelman, S. A. (2000). Preschool children's use of novel predicates to make inductive inferences about people. *Cognitive Development, 15*(3), 263-280.

Keil, F. C. (1992). *Concepts, kinds, and cognitive development*. Cambridge, MA: MIT Press.

Khanna, M. M., & Boland, J. E. (2010). Children's use of language context in lexical ambiguity resolution. *The Quarterly Journal of Experimental Psychology, 63*(1), 160-193.

Klein, D. E., & Murphy, G. L. (2001). The representation of polysemous words. *Journal of Memory and Language, 45*(2), 259-282.

Klein, D. E., & Murphy, G. L. (2002). Paper has been my ruin: conceptual relations of polysemous senses. *Journal of Memory and Language, 47*(4), 548-570.

Lopez, A., Gelman, S. A., Gutheil, G., & Smith, E. E. (1992). The Development of Category-Based Induction. *Child development, 63*(5), 1070-1090.

MacWhinney, B. (2000). *The CHILDES Project: Tools for analyzing talk.* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Markman, E. M. (1991). *Categorization and naming in children*. Cambridge, MA: MIT Press.

McGarrigle, J., & Donaldson, M. (1974). Conservation accidents. *Cognition, 3*(4), 341-350.

Murphy, G. (2004). *The big book of concepts*. Cambridge, MA: MIT press.

Morey, R.D., & Rouder, J.N. (2015). *BayesFactor: Computation of Bayes Factors for Common Designs*. R package version 0.9.12-2

Nunberg, G. (1979). The non-uniqueness of semantic solutions: Polysemy. *Linguistics and Philosophy, 3*(2), 143-184.

Osherson, D. N., Smith, E. E., Wilkie, O., Lopez, A., & Shafir, E. (1990). Category-based induction. *Psychological review, 97*(2), 185.

Rabagliati, H., Marcus, G. F., & Pylkkänen, L. (2010). Shifting senses in lexical semantic development. *Cognition, 117*(1), 17-37.

Rabagliati, H., Pylkkänen, L., & Marcus, G. F. (2013). Top-down influence in young children's linguistic ambiguity resolution. *Developmental psychology, 49*(6), 1076.

Rabagliati, H., & Snedeker, J. (2013). The Truth About Chickens and Bats Ambiguity Avoidance Distinguishes Types of Polysemy. *Psychological Science, 24*(7), 1354-1360.

Sloutsky, V. M., & Fisher, A. V. (2004). Induction and categorization in young children: a similarity-based model. *Journal of Experimental Psychology: General, 133*(2), 166.

Sloutsky, V. M., & Fisher, A. V. (2011). The development of categorization. *Psychology of learning and motivation, 54*, 141-166.

Sloutsky, V. M., & Fisher, A. V. (2012). Linguistic labels: conceptual markers or object features? *Journal of Experimental Child Psychology, 111*(1), 65-86.

Sloutsky, V. M., Fisher, A. V., & Kloos, H. (2015). Conceptual influences on induction: A case for a late onset. *Cognitive psychology, 82*, 1-31.

Sloutsky, V. M., Kloos, H., & Fisher, A. V. (2007). When looks are everything Appearance similarity versus kind information in early induction. *Psych. Science, 18*(2), 179-185.

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*(6), 1478.

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*(6), 1695-1709.

Srinivasan, M. (2016). Concepts as explanatory theories: Evidence from word learning and the development of lexical flexibility. In D. Barner & A. Baron (Eds.), *Core Knowledge and Conceptual Change*. Oxford, UK: OUP.

Srinivasan, M., Al-Mughairy, S., Foushee, R., & Barner, D. (2017). Learning language from within: Children use semantic generalizations to infer word meanings. *Cognition*, *159*, 11-24.

Srinivasan, M., & Rabagliati, H. (2015). How concepts and conventions structure the lexicon: Cross-linguistic evidence from polysemy. *Lingua, 157*, 124-152.

Srinivasan, M., & Snedeker, J. (2011). Judging a book by its cover and its contents: The representation of polysemous and homophonous meanings in four-year-old children. *Cognitive psychology, 62*(4), 245-272.

Srinivasan, M., & Snedeker, J. (2014). Polysemy and the Taxonomic Constraint: Children's Representation of Words that Label Multiple Kinds. *Language Learning and Development, 10*(2), 97-128.

Waxman, S. R., & Lidz, J. L. (2006). Early world learning. *Handbook of child psychology*.

Waxman, S. R., & Markow, D., B. (1995). Words as invitations to form categories: Evidence from 12-to 13-month-old infants. *Cognitive psychology, 29*(3), 257-302.

Xu, F. (2002). The role of language in acquiring object kind concepts in infancy. *Cognition, 85*(3), 223-250.

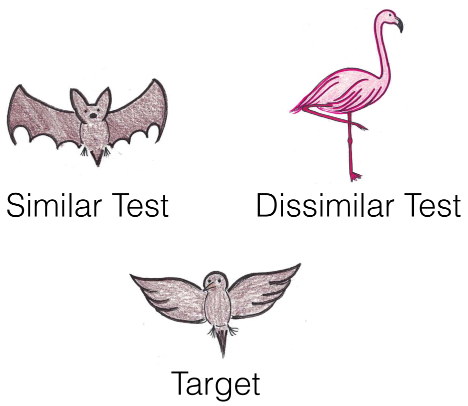

Figure 1. An example stimulus from an inductive reasoning “triad” task. Children are told that *this bat’s legs stay warm at night* (Similar Test picture, top left)*,* and that *this bird’s legs get cold at night* (Dissimilar Test picture, top right), and are then asked which property is held by the bottom *bird* (Target, similar to the bat, same category as the flamingo).

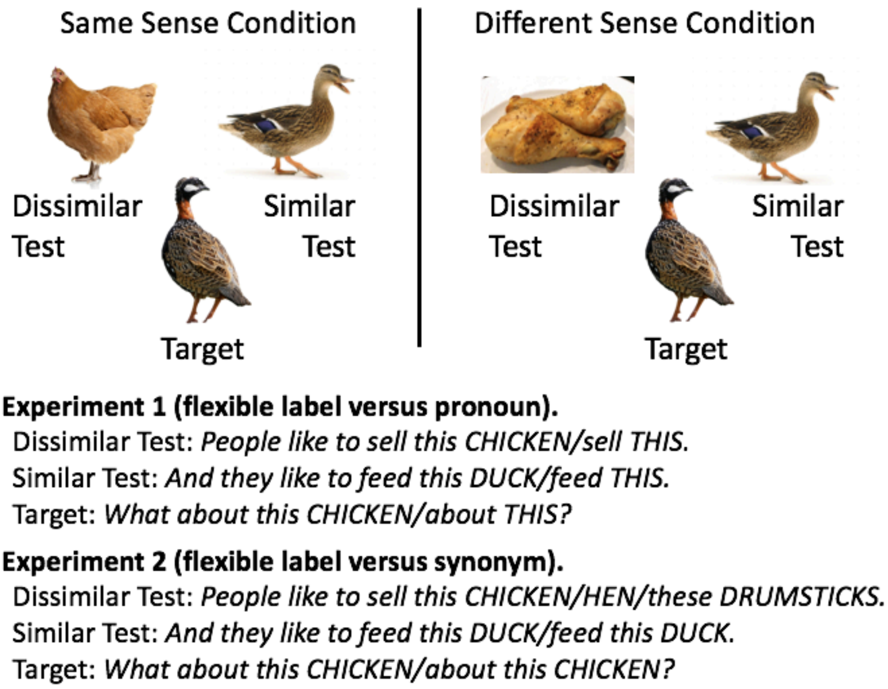

Figure 2. Stimuli and procedure for Experiments 1 and 2.

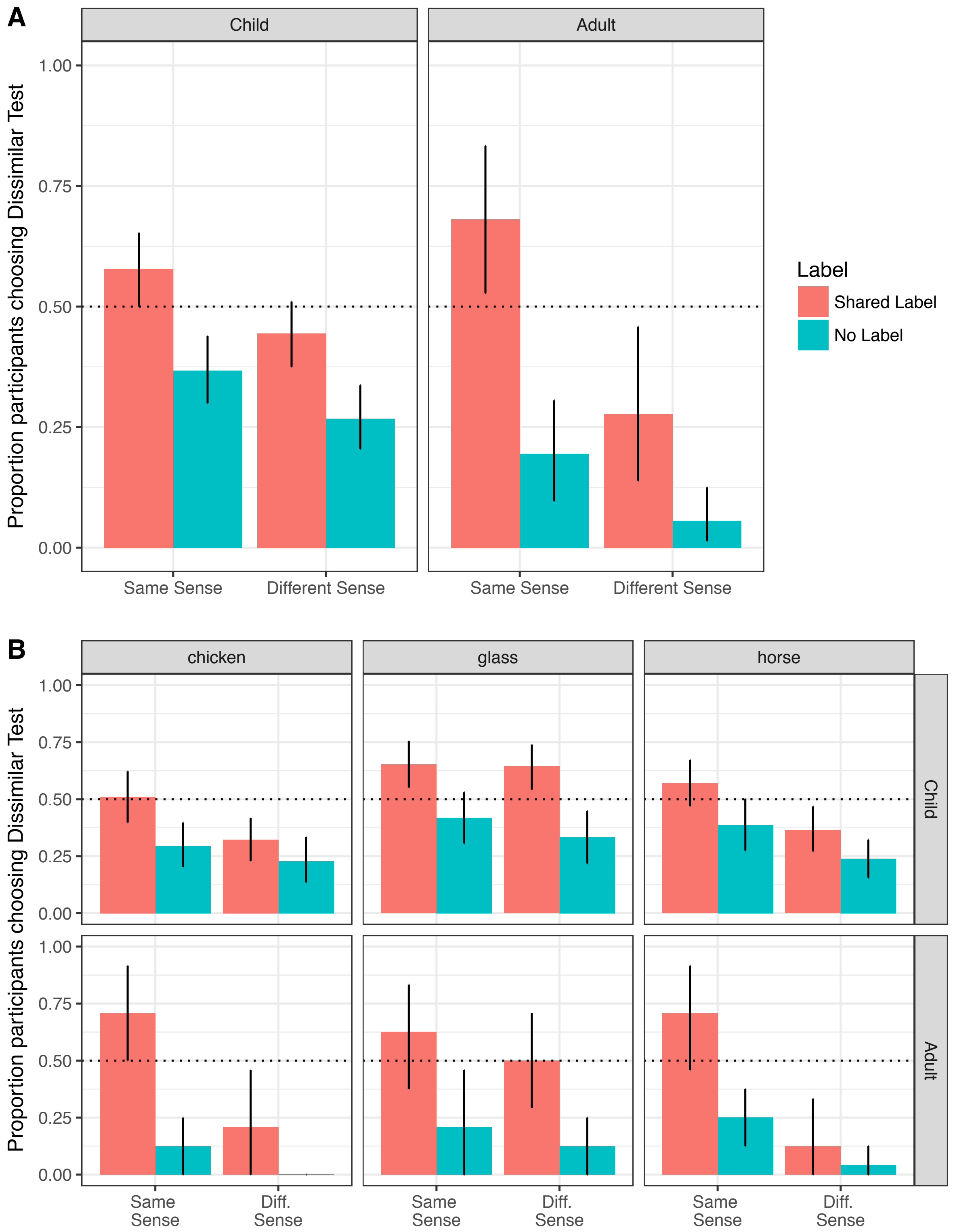

Figure 3. Children's and adults' judgments in Experiment 1 **A.** Overall and **B.** By item. Error bars show 95% confidence intervals around the mean.

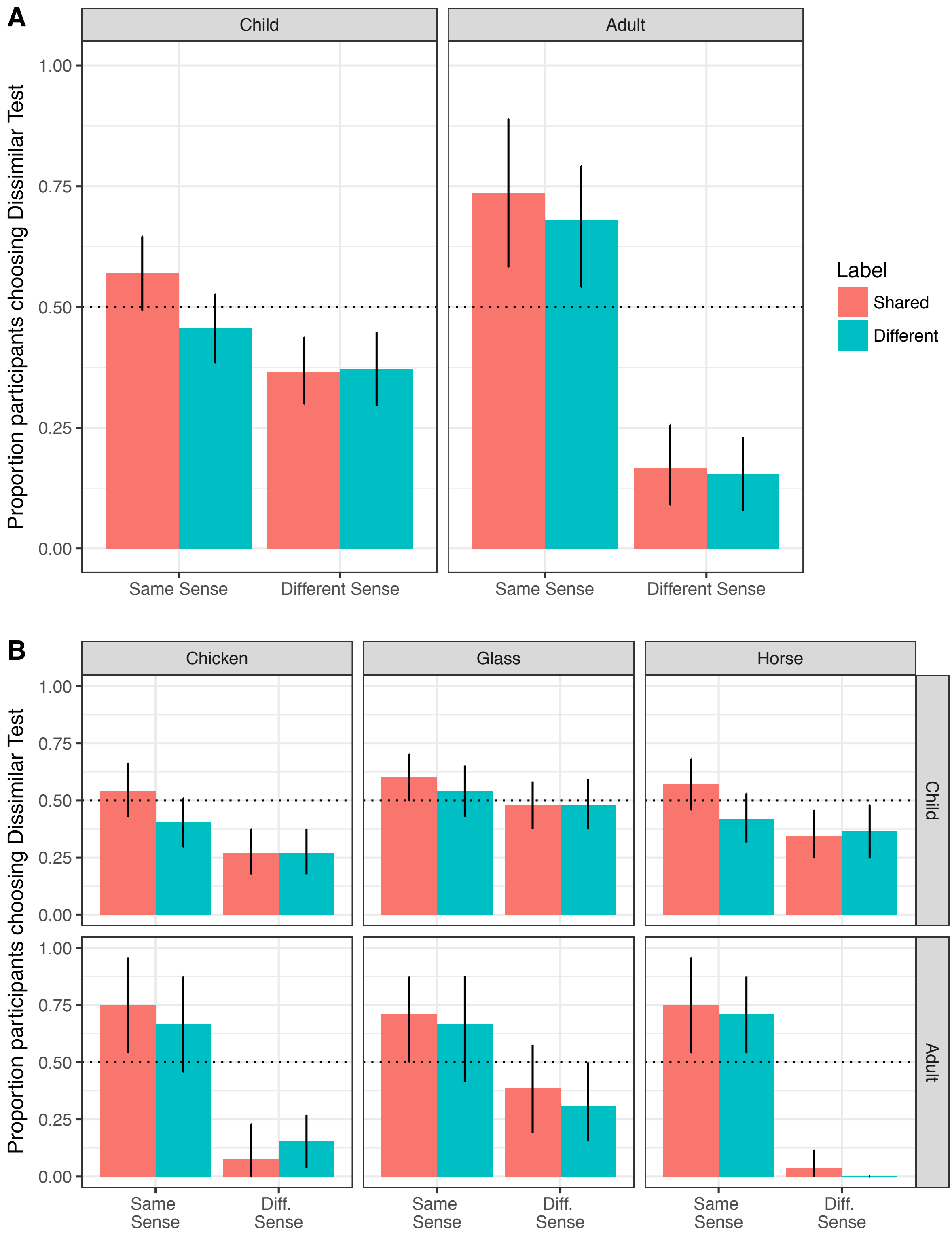

Figure 4. Children's and adults' judgments in Experiment 2 **A.** overall and **B.** by item. Error bars show 95% confidence intervals around the mean.