Triad Polysemy Induction Tasks

Hugh Rabagliati and Mahesh Srinivasan

5 July 2016

**Introduction**

A key debate in developmental psych concerns the role of labels in how children learn and use categories.

Give v brief rundown of broader evidence eg Waxman, Booth

But perhaps the most well articulated debate concerns the role of labels in young children’s reasoning. In particular, children’s inductive inferences have been shown to be preferentially sensitive to labels rather than visual similarity. Argument as to whether this shows use of detailed conceptual and category knowledge when reasoning, or whether it can be explained by a simpler, similarity based model.

The assumption behind these debates is that children rely on labels because labels are robust category markers, eg whether this is for statistical or conceptual reasons.

But that's not quite right: Labels are not robust category markers, in that almost all labels are used to refer to multiple different categories. We call this phenomenon lexical flexibility

Here we investigate what the presence of lex flex means for the role of labels in how children reason about concepts and categories.

**Gelman Sloutsky debate**

overview and history of gelman sloutsky debate. This should be a couple of paragraphs. Begin with Gelman and Markman, and introduce the triad method. Note that folks like Sloutsky have suggested that homonyms are relatively rare – this is true, but other forms of ambiguity are v common.

Perhaps end with final experiment in Sloutsky Fisher JECP, where they show that even phonological similarity between words can affect children's induction, and that the size of this effect is not mediated by similarity of the two concepts under comparison.

It should be clear that flexible labels *could* potentially affect induction. But how, normatively, *should* flexible labels affect induction (e.g., in adults)?

If induction proceeds via reasoning about concepts and world knowledge, then, in fact, lex flex should not have an effect on how people reason. In particular, the plausibility of a conclusion should be based on world knowledge, not any shared label. The use of world knowledge, rather than labels, for induction can be seen in the two arguments below.

*Bears have very heavy bones, therefore, [toy] bears have very heavy bones.*

*Bears have no tails, therefore, [toy] bears have no tails.*

Intuitively the first argument seems implausible, because we know that stuffed bears do not have bones. The second argument, by contrast, seems plausible, because we also know that stuffed bears are supposed to resemble real bears. That is to say, as adults, we don’t get lulled into making odd inferences by the presence of lex flex.

We can use the fact that reasoning based on lexical flexibility relies on world knowledge, not labels, to make clear predictions about how children and adults should behave on this task, based on the two key theories of inductive inference.

Under Gelman’s account, children reason like adults: to the extent that their background knowledge allows, they should make adult-like inductive inferences in the presence of lex flex. Here’s a quote from Gelman and Kalish (2006, Handbook of Child Psych) that illustrates this:

*“However, these findings do not unambiguously locate the source of the labeling effect. 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.” When learning novel words, children do not automatically assume that the words promote inferences, if perceptual cues compete (S. A. Gelman, 2003).”*

But under Sloutsky’s account, children might be lulled into making incorrect inferences. Because, for Sloutsky, children’s inductive reasoning proceeds by similarity, and because labels contribute to similarity, overlapping labels should cause children to perceive that distinct kinds are more similar than they actually are, and make non-adult-like inductive inferences. Adults, by contrast, should not be affected by this, as they can use conceptual knowledge for reasoning.

Quote from Sloutsky Lo (1999) about how children learn the role of labels. Specific focus on labels as a good cue.

*It seems plausible that this developmental shift may be a function of the development of (1) a domain- general record of probabilities, and (2) a domain-specific biological knowledge. The record of probabilities suggests that two remotely similar objects sharing a label are more likely to be considered members of the same category than to be considered members of different categories. This probability could be estimated by its complement, or the probability that two remotely sim- ilar entities that have the same label would be consid- ered members of different categories. This probability of the complement could be estimated by the base rate of homonyms (and homophones), and, therefore, would be negligibly small: it is difficult to come up with a noun having a homonym within the same on- tological class (e.g., living creature having a homonym that indicates a completely different living creature). Hence, as children grow, they acquire more evidence confirming that remotely similar entities sharing the label are likely to be members of the same category. In addition, biological knowledge acquired at school may further corroborate these probabilistic intuitions, suggesting that a category label (e.g., mammal) is the single best predictor of most anatomical and physio- logical properties of mammalian animals.*

Also from Sloutsky & Lo (1999), about how labels should affect similarity judgments for children, but not adults.

*Interestingly enough, the predictive value of labels for induction is independent of similarity: Sloutsky & Lo (1999) demonstrated that shared labels made virtually no contribution to similarity judgments of preadolescents (but made a sizable, yet quantitative, contribution to similarity judgments of young chil- dren). In other words, labels affect both similarity and induction in young children, but affect only induction in preadolescents and adults. Therefore, it seems rea- sonable to conclude that in young children (1) similar- ity and induction are related, and (2) labels contribute quantitatively to both similarity and induction. At the same time, in older children and adults (1) similarity and induction are independent, and (2) labels make no contribution to similarity, but contribute qualita- tively to their inductio*

**Our approach**

Because these theories make different predictions, lexical flexibility promises to inform our understanding of children’s conceptual development. To investigate how children’s and adults’ inferences are affected by lexical flexibility, we used the classic Gelman and Markman “triad” inductive reasoning task. On each trial of this task, participants see three pictures: Test picture A, Test picture B, and a Target picture. Participants are told that Test A has one property, and that Test B had a different property, and then they judge whether the Target has the same property as Test A or Test B. Critically, one of the Test pictures looks more like the Target than the other Test picture. However, in the classic version of this task, the Target picture and the dissimilar Test picture are given the same label. For example, participants might see the triad in Figure X, and be told that the Target picture is a chicken, that Test A is a duck, and that Test B is a chicken (even though Test A looks more like the Target). The participant might then be told that the duck (Test A) has one property (e.g., people keep this duck cold), and the second chicken (TEST B) has another property (e.g., people keep this chicken warm), and will be asked whether the Target shares the property with Test A or Test B. Both children and adults tend to respond according to the label, rather than visual similarity.

We compared the classic version of this task to a “flexibility” version, in which the dissimilar Test picture (Test B) was replaced. The new Test picture was from a distinct category but was thematically related via lexical flexibility; for example, the animal chicken might be replaced with chicken meat. We used the same properties in the classic and flexible version of the experiment (e.g., people keep this duck cold, people keep this chicken warm). These properties were specifically designed so that participants would be less willing to generalize thematically than taxonomically. For example, whereas we expected participants to generalize between chickens in the classic taxonomic experiment, we expected them to not generalize between chickens in the flexible thematic experiment.

In Experiment 1, we tested how the inferences of adults and children varied between these two tasks. To test the role of labels, we contrasted a version in which pictures were given category-marking labels (e.g., chicken and duck) to a version in which labels were replaced with demonstrative pronouns (this one). If children reason based on concepts from the start then, like adults, they should show an effect of Label in the classic taxonomic condition, but should not show an effect of Label in the flexible thematic condition. But if children’s use of labels is quite different from adults, then they should show an effect of Label not only in the taxonomic condition, but also the thematic condition.

**Experiment 1**

**Methods**

**Participants**

Our final child sample was 96 3- and 4-year-olds (48 per age group). 54 children were tested in Edinburgh, at local preschools and in the Edinburgh Developmental Lab, and 42 children were tested in the San Francisco Bay Area, at local museums, preschools and in the Berkeley Developmental Lab. YY children were also excluded (WHY?). An additional XX (right now = 14, final = 24?) adults were also tested, 10 at the University of Edinburgh and 14 at U.C. Berkeley.

**Materials**

Materials were 12 pairs of triads of pictures, and associated properties. For each pair, the Target picture (e.g., a chicken animal) and Test A (e.g., a duck) were always the same, but the picture of Test B varied. Either the Target and Test B illustrated a single sense of a flexible word (e.g., both depicted chicken animals) or they illustrated two thematically related senses of that flexible word (e.g., the Target was a chicken animal, and Test B illustrated chicken meat).

The 12 triads were created based on three different flexible words: Chicken (Target: Animal, Test B: Animal/Meat), Glass (Target: Material (RIGHT?), Test B: Material/Drinking glass), and Horse (Target: Animal, Test B: Animal/Toy horse). 4 different triads, with different pictures and properties, were created for each flexible word. Properties were specifically chosen such that the authors had the intuition that participants should prefer to generalize from Target to Test B taxonomically (e.g, from chicken animal to chicken animal) but not thematically (from chicken animal to chicken meat).

**Procedure**

In order to make the task more interesting, and in order to ensure that participants were more likely to base their judgments on inductive inferences than particular prior knowledge, the task took the form of a game, in which participants had to guess the answer to questions about objects found on the planet Jupp. Participants were told that Jupp is a planet quite like earth, but also a little different.

On each trial, the triad was displayed to the participant on a laptop computer screen, and the experimenter read the properties aloud from a script. The script always took the same form. For instance, on *chicken* trials the experimenter might say

*Do you see this chicken (point to Target)? And do you see this duck (point to Test A)? And do you see this chicken (point to Test B)? On planet Jupp, people keep this duck cold, and they keep this chicken warm. What do they do with this chicken? Do they keep it cold like this duck, or do they keep it warm like this chicken?*

Children were expected to point to, or otherwise indicate, whether the Target had Test A’s property or Test B’s.

We varied whether participants saw classic “taxonomic” induction trials, or flexible “thematic” induction trials between subjects. We varied whether participants heard full labels (eg. *do you see* *this* *chicken?*) or just demonstrative pronouns (*do you see this?*) within subjects. Order of mention of Test A and Test B, and position of the pictures on the page, were counterbalanced between lists. Participants received trials in one of two random orders.

Participants completed two simple taxonomic trials as a warm up, before the 12 test trials.

**Analysis**

We analyzed the proportion of trials on which participants matched the Target to Test B, the picture that was related either taxonomically or thematically. We used mixed effects logistic regressions to model the choices of children and adults. Our dependent variable was participants’ choice (related/unrelated) and our predictor variables were Induction Task (classic taxonomic/flexible thematic), Label Presence (label/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 items.

**Results**

If induction tasks are solved by reasoning about concepts, then participants should show effects of label in the taxonomic condition, when the label indicates that the dissimilar item is in fact from the same taxonomic category as the Target, but should not show effects of label in the flexible thematic condition, as participants realize that it is not indicative of a taxonomic relationship. By contrast, if children use simple similarity to complete these tasks, then we would expect the presence of a shared label to affect their reasoning in both the classic taxonomic and flexible thematic versions of the induction task.

Children’s answers did indeed show effects of Label Presence in both versions of the induction task. As Figure Xa shows, participants were overall more likely to choose the related test picture when it shared a label with the target (STAT). Follow-up analyses showed that this result held in both the taxonomic and thematic conditions (STAT). In addition, children were more likely to choose the related item in the taxonomic task, indicating that THEY ARE NOT STUPID. However, the interaction between label presence and task was not significant. Visual inspection of the results for the three different types of triad (chicken, glass, horse, Fig Xb) shows that the effect of label was consistent across items.

These results are consistent with the idea that children’s induction is not based on conceptual reasoning. However, the data from adults showed a surprising pattern that, potentially, casts doubt on this initial conclusion. In particular, as can be seen in Figure Xc, adults also showed an effect of label (STAT). This effect was somewhat qualified by an interaction between label and task (STAT), and by an effect of task (STAT), but critically the effect of label was present for both the taxonomic and the thematic tasks (although it was smaller in the latter). The presence of the label effect is unexpected in the thematic task, given that adults are assumed – by all theories – to reason inductively using conceptual knowledge.

Fig. 1. Children’s judgments

Child_NoLabel.pdf

Fig. 2. Children’s judgments by items

Child_NoLabel_Items.pdf

Fig. 3. Adult’s judgments

Adult_NoLabel.pdf

Fig. 4. Adult’s judgments by items

Adult_NoLabel_Items.pdf

**Discussion**

In Experiment 1, we showed that children follow a label when reasoning inductively, even generalizing properties across categories that are only related taxonomically. We had assumed that behavior like this would be non-adult-like but, in fact, we found that adults also showed evidence of this effect: they showed a reliable tendency to follow the label even on flexible, thematically-related trials.

Why did adults follow the label on the flexible thematic induction task, and what does this mean for children’s performance? One possibility is that both adults and children complete these inductions tasks using similarity-based reasoning. But this seems unlikely, given that previous work has documented that adults’ inductive reasoning is conceptually driven. Might some third factor be driving the effect of labels in both children and adults in this experiment?

One potential third factor could be pragmatic reasoning. In particular, an unusual feature of the study – from the participants’ point of view – was that the experimenter only provided labels for the pictures on half of the trials. A reasonable inference that participants might draw, is that the experimenter particularly wanted the participant to attend to the label on those trials. This could cause the participant to follow the label when making their inferences, on the assumption that this is what the experimenter wants them to do. In Experiment 2, we test whether label effects are still seen when this pragmatic inference cannot be drawn.

**Experiment 2:**

We reasoned that, in Experiment 1, the absence of labels on some trials might cause participants to follow the labels more closely when they are present. To countermand this effect in Experiment 2, we used labels on all trials, and varied whether labels were shared (two things called *chicken*), or whether labels were synonyms (*chicken* and *rooster/*drumsticks). If previous results reflect sloutsky fisher, should still get effect. If previous results reflect Gelman plus Pragmatics, then should not get effect.

Here, we compare inferences from e.g., Chicken Animal to Duck, and either another Chicken Animal or Chicken Meat. We also vary whether the same label is used twice (both pictures called chicken) or whether a synonym is used (one called chicken, one called drumsticks).

Fig. 5. Children’s judgments

Child_Syn.pdf

Fig. 6. Children’s judgments by items

Child_Syn_Items.pdf

Fig. 7. Adult’s judgments

Adult_Syn.pdf

Fig. 8. Adult’s judgments by items

Adult_Syn_Items.pdf

**Comparison of Experiments 1 and 2**

**Fig. 9. Children’s judgments**

**Child_CompareExpts.pdf**

**Fig. 10. Adult’s judgments**

**Adult_CompareExpts.pdf**

**Fig. 11. Children’s judgments by Experiment and Age**

Child_CompareExpts_Ages.pdf

**Discussion**

No robust effect of synonyms, and effect was signif different between two experiments.

This suggests that the effect in the Experiment 1 was driven by pragmatic reasoning, rather than because shared labels lead to increased perceived similarity.

**Gen discussion**

What can lex flex tell us about role of labels in induction?

What can induction tell us about children's early word meanings?

Our data is inconsistent with both accounts. Sloutsky account needs a better story about label similarity, because can’t explain Expt 2. But Gelman account doesn’t work for Expt 1, as doesn’t have a role for pragmatics.