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, Vouloumanos

But perhaps the most well articulated debate is gelman sloutsky. Two line summary

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. 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 affect. Reasoning should be based on world knowledge, to see whether a conclusion is plausible. Our reliance on world knowledge, rather than labels, for induction can be seen in the two arguments below.

Bears have very heavy bones, therefore, [teddy] bears have very heavy bones.

Bears have pointy ears, therefore, [teddy] bears have pointy ears.

Intuitively the first argument seems less plausible than the second, because we know that stuffed bears do not have bones, and 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 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.

But under Sloutsky’s account, children might be lulled into making incorrect inferences. Because, for Sloutsky, 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.

**Our approach**

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 ha 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 PROPERTY, and the second chicken (TEST B) has OTHER PROPERTY, 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., this duck has PROPERTY, this chicken has PROPERTY). 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 experiment, we expected them to not generalize between chickens in the flexible 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**

**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

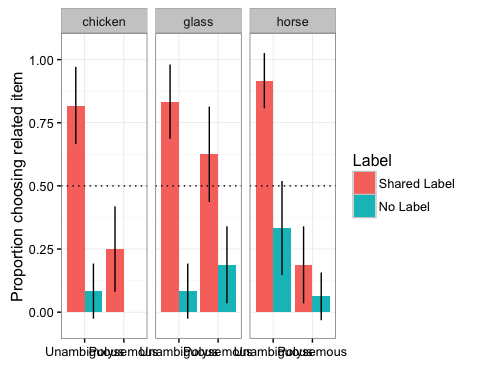
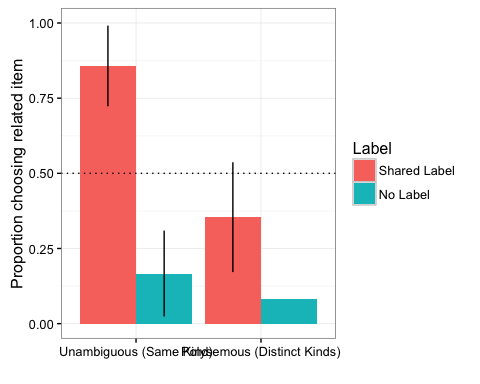
*Do you see this chicken (point to Target)? And do you see this duck (point to Test A)? And do you see this chicken (point to Test B)? On planet Jupp, people keep this duck cold, and they keep this chicken warm. What do they do with this chicken? Do they keep it cold like this duck, or do they keep it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

**Adults**



Label \* Meaning mixed effects model, followed by t-tests against chance for each condition.

Key results:

Effects of both label and meaning. An interaction, too (Note that the regression below failed to converge, but when we drop the random by participant effect of Label, it does converge.)

There is an effect of label for both flexible and unambiguous conditions, but the effect is smaller for flexible.

## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control  
## $checkConv, : Model failed to converge with max|grad| = 0.027305 (tol =  
## 0.001, component 1)

## Generalized linear mixed model fit by maximum likelihood (Laplace  
## Approximation) [glmerMod]  
## Family: binomial ( logit )  
## Formula: Choice ~ Label \* Meaning + (1 + Label | ID) + (1 + Meaning |   
## question.number)  
## Data: Adult  
##   
## AIC BIC logLik deviance df.resid   
## 154.5 185.7 -67.3 134.5 157   
##   
## Scaled residuals:   
## Min 1Q Median 3Q Max   
## -2.3044 -0.3024 -0.1391 0.1809 2.3637   
##   
## Random effects:  
## Groups Name Variance Std.Dev. Corr   
## ID (Intercept) 0.7957 0.892   
## Label.L 3.2488 1.802 -0.30  
## question.number (Intercept) 1.2834 1.133   
## Meaning.L 5.4857 2.342 -0.72  
## Number of obs: 167, groups: ID, 14; question.number, 12  
##   
## Fixed effects:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -0.983598 0.002712 -362.7 <2e-16 \*\*\*  
## Label.L -3.408398 0.002712 -1256.8 <2e-16 \*\*\*  
## Meaning.L -1.740383 0.002801 -621.4 <2e-16 \*\*\*  
## Label.L:Meaning.L 2.356227 0.002801 841.3 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Correlation of Fixed Effects:  
## (Intr) Labl.L Mnng.L  
## Label.L 0.000   
## Meaning.L 0.000 0.000   
## Lbl.L:Mnn.L 0.000 0.000 0.000   
## convergence code: 0  
## Model failed to converge with max|grad| = 0.027305 (tol = 0.001, component 1)

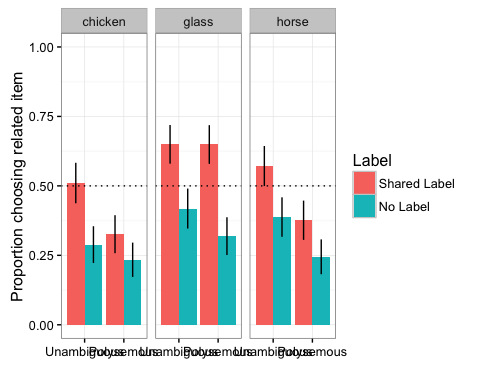
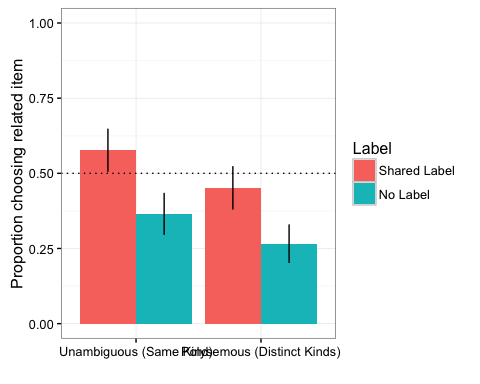
##   
## One Sample t-test  
##   
## data: subset(Adult.Sum, Meaning == "Unambiguous (Same Kind)" & Label == "Shared Label")$Choice  
## t = 4.4907, df = 4, p-value = 0.0109  
## alternative hypothesis: true mean is not equal to 0.5  
## 95 percent confidence interval:  
## 0.6399709 1.0933625  
## sample estimates:  
## mean of x   
## 0.8666667 **NOTE THAT THIS ISN’T QUITE RIGHT: df is wrong. Don’t report this stat, look at the RMarkdown.**

##   
## One Sample t-test  
##   
## data: subset(Adult.Sum, Meaning == "Unambiguous (Same Kind)" & Label == "No Label")$Choice  
## t = -4.4721, df = 5, p-value = 0.006566  
## alternative hypothesis: true mean is not equal to 0.5  
## 95 percent confidence interval:  
## -0.02493319 0.35826652  
## sample estimates:  
## mean of x   
## 0.1666667

##   
## One Sample t-test  
##   
## data: subset(Adult.Sum, Meaning == "Polysemous (Distinct Kinds)" & Label == "Shared Label")$Choice  
## t = -1.3129, df = 7, p-value = 0.2306  
## alternative hypothesis: true mean is not equal to 0.5  
## 95 percent confidence interval:  
## 0.09150396 0.61682938  
## sample estimates:  
## mean of x   
## 0.3541667

##   
## One Sample t-test  
##   
## data: subset(Adult.Sum, Meaning == "Polysemous (Distinct Kinds)" & Label == "No Label")$Choice  
## t = -9.3541, df = 7, p-value = 3.317e-05  
## alternative hypothesis: true mean is not equal to 0.5  
## 95 percent confidence interval:  
## -0.0219954 0.1886621  
## sample estimates:  
## mean of x   
## 0.08333333

Children



Label \* Meaning mixed effects model, followed by t-tests against chance for each condition.

Key results:

Effects of both label and meaning. No interaction

## Generalized linear mixed model fit by maximum likelihood (Laplace  
## Approximation) [glmerMod]  
## Family: binomial ( logit )  
## Formula:   
## Choice ~ Label \* Meaning + (1 + Label | ID) + (1 | question.number)  
## Data: Child  
##   
## AIC BIC logLik deviance df.resid   
## 1464.8 1505.2 -724.4 1448.8 1139   
##   
## Scaled residuals:   
## Min 1Q Median 3Q Max   
## -1.8713 -0.7172 -0.4626 0.8564 2.2659   
##   
## Random effects:  
## Groups Name Variance Std.Dev. Corr  
## ID (Intercept) 0.2197 0.4688   
## Label.L 0.9090 0.9534 0.02  
## question.number (Intercept) 0.1173 0.3425   
## Number of obs: 1147, groups: ID, 95; question.number, 12  
##   
## Fixed effects:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -0.42689 0.12985 -3.288 0.001011 \*\*   
## Label.L -0.68400 0.13836 -4.944 7.67e-07 \*\*\*  
## Meaning.L -0.40265 0.11726 -3.434 0.000595 \*\*\*  
## Label.L:Meaning.L 0.05675 0.19333 0.294 0.769104   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Correlation of Fixed Effects:  
## (Intr) Labl.L Mnng.L  
## Label.L 0.051   
## Meaning.L 0.037 0.035   
## Lbl.L:Mnn.L 0.021 0.048 0.043

##   
## One Sample t-test  
##   
## data: subset(Child.Sum, Meaning == "Unambiguous (Same Kind)" & Label == "Shared Label")$Choice  
## t = 1.995, df = 48, p-value = 0.05174  
## alternative hypothesis: true mean is not equal to 0.5  
## 95 percent confidence interval:  
## 0.4993860 0.6570766  
## sample estimates:  
## mean of x   
## 0.5782313

##   
## One Sample t-test  
##   
## data: subset(Child.Sum, Meaning == "Unambiguous (Same Kind)" & Label == "No Label")$Choice  
## t = -3.6101, df = 48, p-value = 0.0007292  
## alternative hypothesis: true mean is not equal to 0.5  
## 95 percent confidence interval:  
## 0.2913484 0.4406244  
## sample estimates:  
## mean of x   
## 0.3659864

##   
## One Sample t-test  
##   
## data: subset(Child.Sum, Meaning == "Polysemous (Distinct Kinds)" & Label == "Shared Label")$Choice  
## t = -1.5838, df = 45, p-value = 0.1202  
## alternative hypothesis: true mean is not equal to 0.5  
## 95 percent confidence interval:  
## 0.3707760 0.5154559  
## sample estimates:  
## mean of x   
## 0.4431159

##   
## One Sample t-test  
##   
## data: subset(Child.Sum, Meaning == "Polysemous (Distinct Kinds)" & Label == "No Label")$Choice  
## t = -7.0353, df = 45, p-value = 9.034e-09  
## alternative hypothesis: true mean is not equal to 0.5  
## 95 percent confidence interval:  
## 0.2017307 0.3345011  
## sample estimates:  
## mean of x   
## 0.2681159

Discussion

These results are potentially evidence for Sloutsky account: Hearing a label for a Polysemous category made children more likely to assume that properties transferred between two taxonomically distinct concepts.

But one surprising result that does not make sense under either Sloutsky or Gelman account: Adults go with label on lex flex condition. Why?

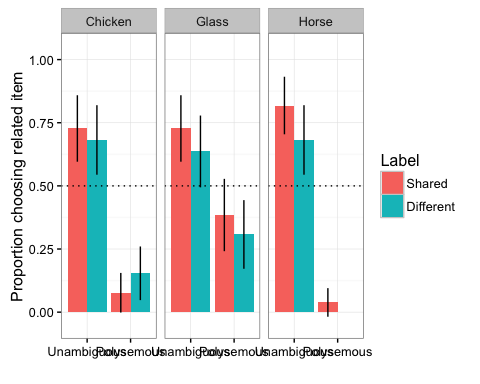
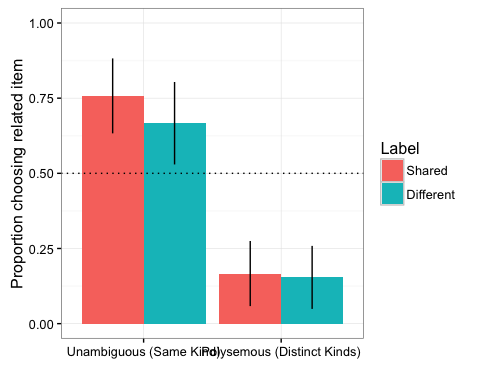
Perhaps there is a pragmatic explanation of participants’ behavior. In particular, participants might follow label on this task because of pragmatic reasoning: Why would the experimenter use these distinct labels if they did not want us to follow those labels during reasoning?

Experiment 2:

Always use labels, but vary if they are shared or synonyms. 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).

Adults



Label \* Meaning mixed effects model, followed by t-tests against chance for each condition.

Key result

Main effect of word type. No effect of label, or interaction.

## Generalized linear mixed model fit by maximum likelihood (Laplace  
## Approximation) [glmerMod]  
## Family: binomial ( logit )  
## Formula:   
## UnrelatedChoice ~ WordType \* LabelType + (1 + LabelType | SubjNo) +   
## (1 | ItemNo)  
## Data: Adult  
##   
## AIC BIC logLik deviance df.resid   
## 289.9 319.2 -137.0 273.9 280   
##   
## Scaled residuals:   
## Min 1Q Median 3Q Max   
## -3.8831 -0.4629 0.2711 0.4453 3.1935   
##   
## Random effects:  
## Groups Name Variance Std.Dev. Corr   
## SubjNo (Intercept) 1.4689 1.2120   
## LabelTypeDifferent 0.1927 0.4389 -1.00  
## ItemNo (Intercept) 0.4879 0.6985   
## Number of obs: 288, groups: SubjNo, 24; ItemNo, 12  
##   
## Fixed effects:  
## Estimate Std. Error z value Pr(>|z|)  
## (Intercept) -1.6063 0.5978 -2.687 0.00721  
## WordTypePolysemous 3.6682 0.8034 4.566 4.98e-06  
## LabelTypeDifferent 0.6668 0.5470 1.219 0.22278  
## WordTypePolysemous:LabelTypeDifferent -0.7464 0.8153 -0.915 0.35995  
##   
## (Intercept) \*\*   
## WordTypePolysemous \*\*\*  
## LabelTypeDifferent   
## WordTypePolysemous:LabelTypeDifferent   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Correlation of Fixed Effects:  
## (Intr) WrdTyP LblTyD  
## WrdTypPlysm -0.729   
## LblTypDffrn -0.708 0.596   
## WrdTypP:LTD 0.544 -0.728 -0.770

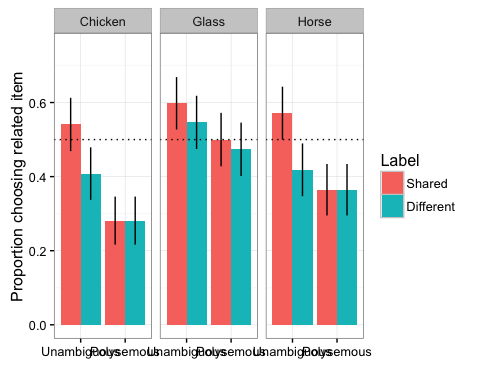
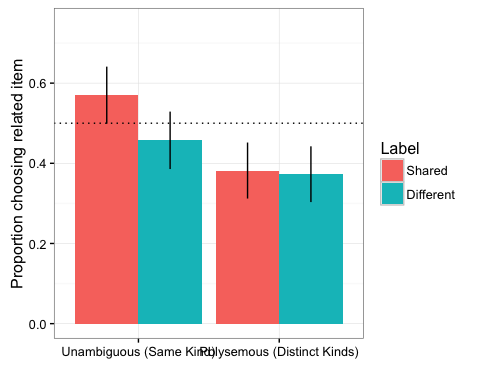
##   
## One Sample t-test  
##   
## data: subset(Adult.Sum, WordType == "Non-Polysemous" & LabelType == "Same")$Choice  
## t = 2.8333, df = 10, p-value = 0.01775  
## alternative hypothesis: true mean is not equal to 0.5  
## 95 percent confidence interval:  
## 0.5550177 0.9601338  
## sample estimates:  
## mean of x   
## 0.7575758

##   
## One Sample t-test  
##   
## data: subset(Adult.Sum, WordType == "Non-Polysemous" & LabelType == "Different")$Choice  
## t = 2.2361, df = 10, p-value = 0.04933  
## alternative hypothesis: true mean is not equal to 0.5  
## 95 percent confidence interval:  
## 0.5005910 0.8327423  
## sample estimates:  
## mean of x   
## 0.6666667

##   
## One Sample t-test  
##   
## data: subset(Adult.Sum, WordType == "Polysemous" & LabelType == "Different")$Choice  
## t = -8.6841, df = 12, p-value = 1.608e-06  
## alternative hypothesis: true mean is not equal to 0.5  
## 95 percent confidence interval:  
## 0.06699738 0.24069493  
## sample estimates:  
## mean of x   
## 0.1538462

##   
## One Sample t-test  
##   
## data: subset(Adult.Sum, WordType == "Polysemous" & LabelType == "Same")$Choice  
## t = -7.2111, df = 12, p-value = 1.07e-05  
## alternative hypothesis: true mean is not equal to 0.5  
## 95 percent confidence interval:  
## 0.06595101 0.26738233  
## sample estimates:  
## mean of x   
## 0.1666667

Children



Label \* Meaning mixed effects model, followed by t-tests against chance for each condition.

Key result

Main effect of word type. No effect of label, or interaction.

## Generalized linear mixed model fit by maximum likelihood (Laplace  
## Approximation) [glmerMod]  
## Family: binomial ( logit )  
## Formula: Choice ~ Meaning \* LabelType + (1 + LabelType | SubjNo) + (1 +   
## Meaning | question.number)  
## Data: Child  
##   
## AIC BIC logLik deviance df.resid   
## 1520.1 1570.6 -750.0 1500.1 1151   
##   
## Scaled residuals:   
## Min 1Q Median 3Q Max   
## -1.8804 -0.7276 -0.4563 0.8167 2.1833   
##   
## Random effects:  
## Groups Name Variance Std.Dev. Corr   
## SubjNo (Intercept) 0.4282 0.6544   
## LabelTypeDifferent 0.4205 0.6485 -0.17  
## question.number (Intercept) 0.1638 0.4048   
## Meaning.L 0.1039 0.3224 0.67   
## Number of obs: 1161, groups: SubjNo, 97; question.number, 12  
##   
## Fixed effects:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -0.2732 0.1501 -1.820 0.06875 .   
## Meaning.L -0.4905 0.1628 -3.013 0.00258 \*\*  
## LabelTypeDifferent -0.1417 0.0939 -1.509 0.13138   
## Meaning.L:LabelTypeDifferent 0.1932 0.1330 1.453 0.14622   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Correlation of Fixed Effects:  
## (Intr) Mnng.L LblTyD  
## Meaning.L 0.316   
## LblTypDffrn -0.057 0.004   
## Mnng.L:LbTD 0.002 -0.085 0.031

##   
## One Sample t-test  
##   
## data: subset(Child.Sum, WordType == "Non-Polysemous" & LabelType == "Same")$Choice  
## t = 1.7633, df = 48, p-value = 0.08422  
## alternative hypothesis: true mean is not equal to 0.5  
## 95 percent confidence interval:  
## 0.4901714 0.6499646  
## sample estimates:  
## mean of x   
## 0.570068

##   
## One Sample t-test  
##   
## data: subset(Child.Sum, WordType == "Non-Polysemous" & LabelType == "Different")$Choice  
## t = -1.1923, df = 48, p-value = 0.239  
## alternative hypothesis: true mean is not equal to 0.5  
## 95 percent confidence interval:  
## 0.3830445 0.5298807  
## sample estimates:  
## mean of x   
## 0.4564626

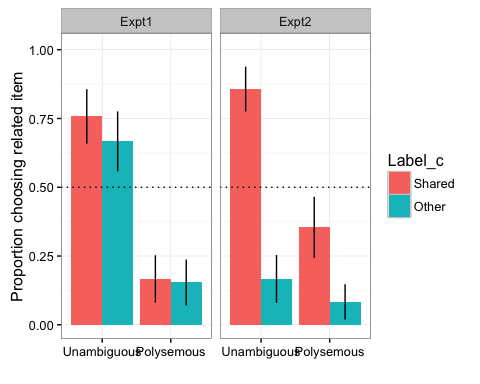
##   
## One Sample t-test  
##   
## data: subset(Child.Sum, WordType == "Polysemous" & LabelType == "Different")$Choice  
## t = -3.3806, df = 47, p-value = 0.001464  
## alternative hypothesis: true mean is not equal to 0.5  
## 95 percent confidence interval:  
## 0.2961843 0.4482601  
## sample estimates:  
## mean of x   
## 0.3722222

##   
## One Sample t-test  
##   
## data: subset(Child.Sum, WordType == "Polysemous" & LabelType == "Same")$Choice  
## t = -3.0723, df = 47, p-value = 0.003528  
## alternative hypothesis: true mean is not equal to 0.5  
## 95 percent confidence interval:  
## 0.3046424 0.4592465  
## sample estimates:  
## mean of x   
## 0.3819444

Comparison of Experiments 1 and 2

Adults

Graph of the two experiments side-by-side. Other = Synonym in Expt 1, No Label in Expt 2



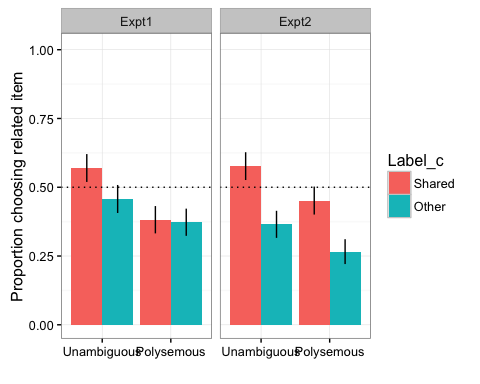
Data analysis. Key finding:

Interaction between effect of label and experiment. No three way interaction between label, condition and experiment.

Suggests that effect of label is really a reaction to hearing “This” used.

## Generalized linear mixed model fit by maximum likelihood (Laplace  
## Approximation) [glmerMod]  
## Family: binomial ( logit )  
## Formula: Choice ~ Meaning \* Label\_c \* Expt + (1 | ID)  
## Data: Adult  
##   
## AIC BIC logLik deviance df.resid   
## 451.5 488.6 -216.8 433.5 446   
##   
## Scaled residuals:   
## Min 1Q Median 3Q Max   
## -2.9732 -0.4793 -0.3441 0.4488 4.5824   
##   
## Random effects:  
## Groups Name Variance Std.Dev.  
## ID (Intercept) 0.5409 0.7354   
## Number of obs: 455, groups: ID, 38  
##   
## Fixed effects:  
## Estimate  
## (Intercept) -0.5725  
## MeaningPolysemous (Distinct Kinds) -1.1484  
## Label\_cOther -0.7776  
## ExptExpt2 -0.2051  
## MeaningPolysemous (Distinct Kinds):Label\_cOther 0.2641  
## MeaningPolysemous (Distinct Kinds):ExptExpt2 0.2812  
## Label\_cOther:ExptExpt2 -0.6234  
## MeaningPolysemous (Distinct Kinds):Label\_cOther:ExptExpt2 0.1600  
## Std. Error  
## (Intercept) 0.1870  
## MeaningPolysemous (Distinct Kinds) 0.1904  
## Label\_cOther 0.1402  
## ExptExpt2 0.1872  
## MeaningPolysemous (Distinct Kinds):Label\_cOther 0.1387  
## MeaningPolysemous (Distinct Kinds):ExptExpt2 0.1870  
## Label\_cOther:ExptExpt2 0.1393  
## MeaningPolysemous (Distinct Kinds):Label\_cOther:ExptExpt2 0.1386  
## z value Pr(>|z|)  
## (Intercept) -3.061 0.00221  
## MeaningPolysemous (Distinct Kinds) -6.033 1.61e-09  
## Label\_cOther -5.546 2.92e-08  
## ExptExpt2 -1.096 0.27309  
## MeaningPolysemous (Distinct Kinds):Label\_cOther 1.905 0.05684  
## MeaningPolysemous (Distinct Kinds):ExptExpt2 1.504 0.13265  
## Label\_cOther:ExptExpt2 -4.474 7.66e-06  
## MeaningPolysemous (Distinct Kinds):Label\_cOther:ExptExpt2 1.155 0.24822  
##   
## (Intercept) \*\*   
## MeaningPolysemous (Distinct Kinds) \*\*\*  
## Label\_cOther \*\*\*  
## ExptExpt2   
## MeaningPolysemous (Distinct Kinds):Label\_cOther .   
## MeaningPolysemous (Distinct Kinds):ExptExpt2   
## Label\_cOther:ExptExpt2 \*\*\*  
## MeaningPolysemous (Distinct Kinds):Label\_cOther:ExptExpt2   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Correlation of Fixed Effects:  
## (Intr) MnP(DK) Lbl\_cO ExptE2 MnP(DK):L\_O MP(DK):E L\_O:EE  
## MnngP(DKnd) -0.054   
## Label\_cOthr 0.095 0.173   
## ExptExpt2 0.332 -0.048 0.109   
## MnP(DK):L\_O 0.145 0.076 -0.056 0.118   
## MnP(DK):EE2 -0.068 0.309 0.110 -0.073 0.101   
## Lbl\_cOt:EE2 0.105 0.139 0.411 0.094 -0.091 0.140   
## MP(DK):L\_O: 0.121 0.097 -0.086 0.148 0.401 0.086 -0.050

Children



Data analysis

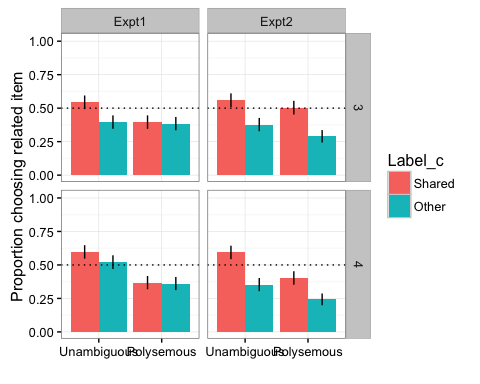
Interaction between effect of label and experiment. No three way interaction between label, condition and experiment.

Suggests that effect of label is really a reaction to hearing “This” used.

## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control  
## $checkConv, : Model failed to converge with max|grad| = 0.00576668 (tol =  
## 0.001, component 1)

## Generalized linear mixed model fit by maximum likelihood (Laplace  
## Approximation) [glmerMod]  
## Family: binomial ( logit )  
## Formula: Choice ~ Meaning \* Label\_c \* Expt + (1 + Label\_c | ID) + (1 +   
## Meaning | QuNum)  
## Data: Child  
##   
## AIC BIC logLik deviance df.resid   
## 2961.3 3041.7 -1466.6 2933.3 2294   
##   
## Scaled residuals:   
## Min 1Q Median 3Q Max   
## -2.0105 -0.7298 -0.4559 0.8280 2.4445   
##   
## Random effects:  
## Groups Name Variance Std.Dev. Corr   
## ID (Intercept) 0.32674 0.5716   
## Label\_cOther 0.44133 0.6643 -0.09  
## QuNum (Intercept) 0.15473 0.3934   
## MeaningPolysemous (Distinct Kinds) 0.04376 0.2092 0.51   
## Number of obs: 2308, groups: ID, 192; QuNum, 12  
##   
## Fixed effects:  
## Estimate  
## (Intercept) -0.35411  
## MeaningPolysemous (Distinct Kinds) -0.31911  
## Label\_cOther -0.32058  
## ExptExpt2 -0.08422  
## MeaningPolysemous (Distinct Kinds):Label\_cOther 0.07873  
## MeaningPolysemous (Distinct Kinds):ExptExpt2 0.01976  
## Label\_cOther:ExptExpt2 -0.17209  
## MeaningPolysemous (Distinct Kinds):Label\_cOther:ExptExpt2 -0.05489  
## Std. Error  
## (Intercept) 0.13005  
## MeaningPolysemous (Distinct Kinds) 0.08736  
## Label\_cOther 0.06793  
## ExptExpt2 0.06283  
## MeaningPolysemous (Distinct Kinds):Label\_cOther 0.06767  
## MeaningPolysemous (Distinct Kinds):ExptExpt2 0.06275  
## Label\_cOther:ExptExpt2 0.06751  
## MeaningPolysemous (Distinct Kinds):Label\_cOther:ExptExpt2 0.06742  
## z value Pr(>|z|)  
## (Intercept) -2.723 0.00647  
## MeaningPolysemous (Distinct Kinds) -3.653 0.00026  
## Label\_cOther -4.719 2.37e-06  
## ExptExpt2 -1.340 0.18010  
## MeaningPolysemous (Distinct Kinds):Label\_cOther 1.163 0.24463  
## MeaningPolysemous (Distinct Kinds):ExptExpt2 0.315 0.75289  
## Label\_cOther:ExptExpt2 -2.549 0.01080  
## MeaningPolysemous (Distinct Kinds):Label\_cOther:ExptExpt2 -0.814 0.41554  
##   
## (Intercept) \*\*   
## MeaningPolysemous (Distinct Kinds) \*\*\*  
## Label\_cOther \*\*\*  
## ExptExpt2   
## MeaningPolysemous (Distinct Kinds):Label\_cOther   
## MeaningPolysemous (Distinct Kinds):ExptExpt2   
## Label\_cOther:ExptExpt2 \*   
## MeaningPolysemous (Distinct Kinds):Label\_cOther:ExptExpt2   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Correlation of Fixed Effects:  
## (Intr) MnP(DK) Lbl\_cO ExptE2 MnP(DK):L\_O MP(DK):E L\_O:EE  
## MnngP(DKnd) 0.323   
## Label\_cOthr -0.006 0.014   
## ExptExpt2 0.014 0.010 0.024   
## MnP(DK):L\_O 0.009 -0.024 0.040 0.012   
## MnP(DK):EE2 0.006 0.016 0.009 0.040 0.019   
## Lbl\_cOt:EE2 0.012 0.010 0.026 -0.028 0.015 0.013   
## MP(DK):L\_O: 0.005 0.014 0.014 0.013 0.019 -0.030 0.037  
## convergence code: 0  
## Model failed to converge with max|grad| = 0.00576668 (tol = 0.001, component 1)

**Analysis by Age: Do**

We can also look at how this varies by age  And analysis (note that model doesn't converge with random intercepts for subjects, just items).

## Generalized linear mixed model fit by maximum likelihood (Laplace  
## Approximation) [glmerMod]  
## Family: binomial ( logit )  
## Formula: Choice ~ Meaning \* Label\_c \* Age \* Expt + (1 | QuNum)  
## Data: Child  
##   
## AIC BIC logLik deviance df.resid   
## 3048.8 3146.5 -1507.4 3014.8 2291   
##   
## Scaled residuals:   
## Min 1Q Median 3Q Max   
## -1.55 -0.83 -0.60 1.01 2.36   
##   
## Random effects:  
## Groups Name Variance Std.Dev.  
## QuNum (Intercept) 0.11 0.33   
## Number of obs: 2308, groups: QuNum, 12  
##   
## Fixed effects:  
## Estimate  
## (Intercept) -0.2996  
## MeaningPolysemous (Distinct Kinds) -0.2674  
## Label\_cOther -0.2803  
## Age4 -0.0069  
## ExptExpt2 -0.0730  
## MeaningPolysemous (Distinct Kinds):Label\_cOther 0.0608  
## MeaningPolysemous (Distinct Kinds):Age4 -0.1035  
## Label\_cOther:Age4 0.0135  
## MeaningPolysemous (Distinct Kinds):ExptExpt2 0.0198  
## Label\_cOther:ExptExpt2 -0.1509  
## Age4:ExptExpt2 -0.0709  
## MeaningPolysemous (Distinct Kinds):Label\_cOther:Age4 0.0116  
## MeaningPolysemous (Distinct Kinds):Label\_cOther:ExptExpt2 -0.0501  
## MeaningPolysemous (Distinct Kinds):Age4:ExptExpt2 0.0171  
## Label\_cOther:Age4:ExptExpt2 -0.0233  
## MeaningPolysemous (Distinct Kinds):Label\_cOther:Age4:ExptExpt2 0.0390  
## Std. Error  
## (Intercept) 0.1054  
## MeaningPolysemous (Distinct Kinds) 0.0437  
## Label\_cOther 0.0438  
## Age4 0.0437  
## ExptExpt2 0.0437  
## MeaningPolysemous (Distinct Kinds):Label\_cOther 0.0437  
## MeaningPolysemous (Distinct Kinds):Age4 0.0437  
## Label\_cOther:Age4 0.0437  
## MeaningPolysemous (Distinct Kinds):ExptExpt2 0.0437  
## Label\_cOther:ExptExpt2 0.0437  
## Age4:ExptExpt2 0.0437  
## MeaningPolysemous (Distinct Kinds):Label\_cOther:Age4 0.0438  
## MeaningPolysemous (Distinct Kinds):Label\_cOther:ExptExpt2 0.0437  
## MeaningPolysemous (Distinct Kinds):Age4:ExptExpt2 0.0437  
## Label\_cOther:Age4:ExptExpt2 0.0437  
## MeaningPolysemous (Distinct Kinds):Label\_cOther:Age4:ExptExpt2 0.0438  
## z value  
## (Intercept) -2.8  
## MeaningPolysemous (Distinct Kinds) -6.1  
## Label\_cOther -6.4  
## Age4 -0.2  
## ExptExpt2 -1.7  
## MeaningPolysemous (Distinct Kinds):Label\_cOther 1.4  
## MeaningPolysemous (Distinct Kinds):Age4 -2.4  
## Label\_cOther:Age4 0.3  
## MeaningPolysemous (Distinct Kinds):ExptExpt2 0.5  
## Label\_cOther:ExptExpt2 -3.4  
## Age4:ExptExpt2 -1.6  
## MeaningPolysemous (Distinct Kinds):Label\_cOther:Age4 0.3  
## MeaningPolysemous (Distinct Kinds):Label\_cOther:ExptExpt2 -1.1  
## MeaningPolysemous (Distinct Kinds):Age4:ExptExpt2 0.4  
## Label\_cOther:Age4:ExptExpt2 -0.5  
## MeaningPolysemous (Distinct Kinds):Label\_cOther:Age4:ExptExpt2 0.9  
## Pr(>|z|)  
## (Intercept) 0.004  
## MeaningPolysemous (Distinct Kinds) 1e-09  
## Label\_cOther 2e-10  
## Age4 0.874  
## ExptExpt2 0.095  
## MeaningPolysemous (Distinct Kinds):Label\_cOther 0.164  
## MeaningPolysemous (Distinct Kinds):Age4 0.018  
## Label\_cOther:Age4 0.757  
## MeaningPolysemous (Distinct Kinds):ExptExpt2 0.651  
## Label\_cOther:ExptExpt2 6e-04  
## Age4:ExptExpt2 0.105  
## MeaningPolysemous (Distinct Kinds):Label\_cOther:Age4 0.791  
## MeaningPolysemous (Distinct Kinds):Label\_cOther:ExptExpt2 0.251  
## MeaningPolysemous (Distinct Kinds):Age4:ExptExpt2 0.695  
## Label\_cOther:Age4:ExptExpt2 0.593  
## MeaningPolysemous (Distinct Kinds):Label\_cOther:Age4:ExptExpt2 0.374  
##   
## (Intercept) \*\*   
## MeaningPolysemous (Distinct Kinds) \*\*\*  
## Label\_cOther \*\*\*  
## Age4   
## ExptExpt2 .   
## MeaningPolysemous (Distinct Kinds):Label\_cOther   
## MeaningPolysemous (Distinct Kinds):Age4 \*   
## Label\_cOther:Age4   
## MeaningPolysemous (Distinct Kinds):ExptExpt2   
## Label\_cOther:ExptExpt2 \*\*\*  
## Age4:ExptExpt2   
## MeaningPolysemous (Distinct Kinds):Label\_cOther:Age4   
## MeaningPolysemous (Distinct Kinds):Label\_cOther:ExptExpt2   
## MeaningPolysemous (Distinct Kinds):Age4:ExptExpt2   
## Label\_cOther:Age4:ExptExpt2   
## MeaningPolysemous (Distinct Kinds):Label\_cOther:Age4:ExptExpt2   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

##   
## Correlation matrix not shown by default, as p = 16 > 12.  
## Use print(...., correlation=TRUE) or  
## vcov(....) if you need it

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?