

The Power of the Prior:

Asymmetries in word and word class learning

Annie Gagliardi & Jeff Lidz
University of Maryland



Questions

- What role do prior hypotheses about word meanings play in the word learning process?
- What about different priors for word meanings and word classes?

Background

Xu & Tenenbaum (2007) examined novel noun learning. A puppet labeled three Dalmatians in an array of animals and other objects as *blicks*. Children tended to think *blick* meant *Dalmatian* rather than *dog* or *animal*. This behavior was attributed to the **likelihood**. The **size principle** caused learners to choose the smallest hypothesis consistent with the data.

Xu & Tenenbaum modeled this using Bayes' Rule:

$$\text{posterior probability of hypothesis} \rightarrow p(h | D) = \frac{\text{likelihood: probability of data given the hypothesis } p(D | h) \times \text{prior probability of hypothesis } p(h)}{\sum_h p(D | h) p(h)}$$

Experiments

Experiment 1: Nouns and adjectives

Children were taught two **nouns** or two **adjectives**. During training a puppet would point out three exemplars for each word from an array of animals and vehicles, and say 'Look it's a **blick!**' or 'Look, it's a **blicky one!**'. At test children were shown a new array and asked to find more **blicks** or more **blicky ones**. Table 1 shows the linguistic stimuli, a sample array can be seen in Figure 1.

Table 1: Linguistic stimuli used in Exp. 1

| Condition | | |
|------------------|--------|-------|
| Noun | blick | fep |
| Adjective | blicky | fepfy |

Figure 1: Sample training/test array



Experiment 2: Nouns, Adjectives and Noun Classes

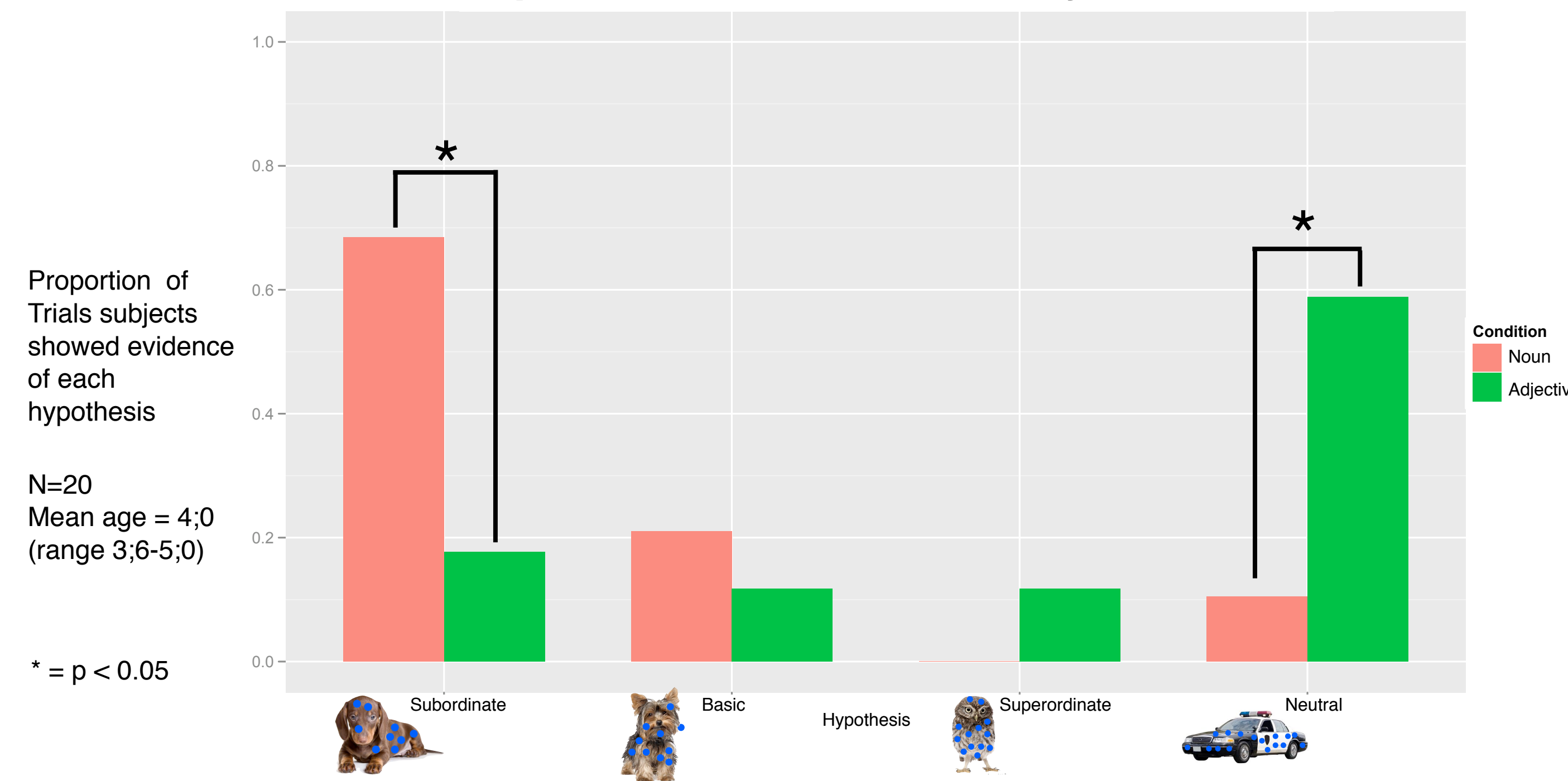
The procedure was identical to Experiment 1, except that children were taught either four **nouns**, four **adjectives** or two adjectival roots and two suffixes, indicating **noun class**. The arrays were identical to those used in Exp. 1 (Figure 1), and the linguistic stimuli can be seen in Table 2.

Table 2: Linguistic stimuli used in Exp. 2

| Condition | | | | |
|-------------------|----------|--------|----------|--------|
| Noun | blick | fep | dax | piff |
| Adjective | blicky | fepfy | daxy | piffy |
| Noun Class | blick-sa | fep-sa | blick-do | fep-do |

Results

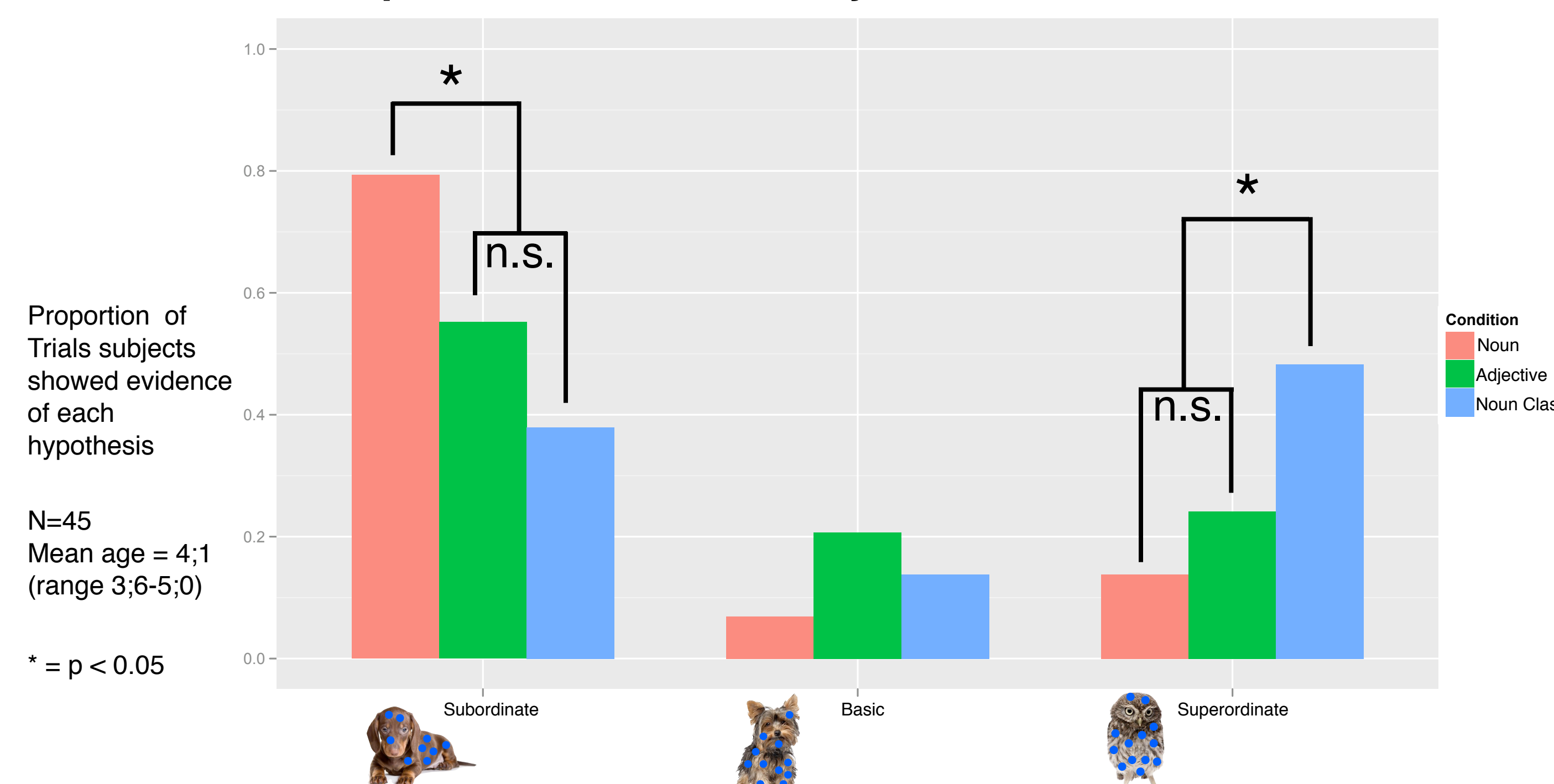
Experiment 1: Nouns and Adjectives



Children learning **nouns** behaved similarly to those in the Xu and Tenenbaum experiment, preferring to pick the smallest hypothesis consistent with the data.

Children learning **adjectives** entertained different hypotheses, learning adjective meanings independent of the noun modified by the adjective.

Experiment 2: Nouns, Adjectives and Noun Classes



Children learning **nouns** behaved similarly to those in Experiment 1

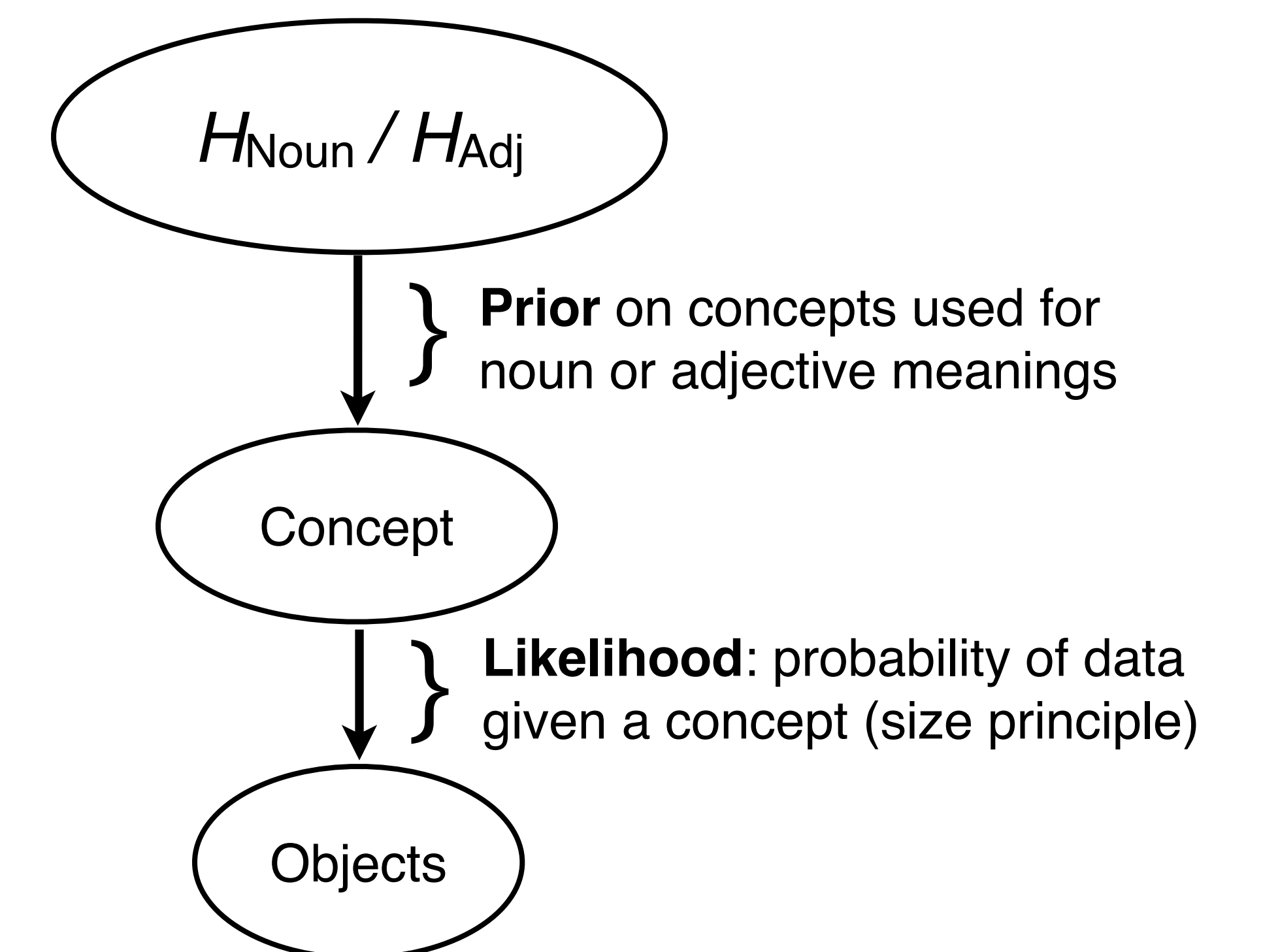
Children learning **adjectives** now show a preference for adjective meanings that take into account the noun being modified.

Children learning **noun classes** show a distinct pattern, splitting their preferences to pick the smallest or largest set consistent with the data.

Model

Below we sketch the outline of a model - implementation is in progress

Nouns and Adjectives (Exp. 1)



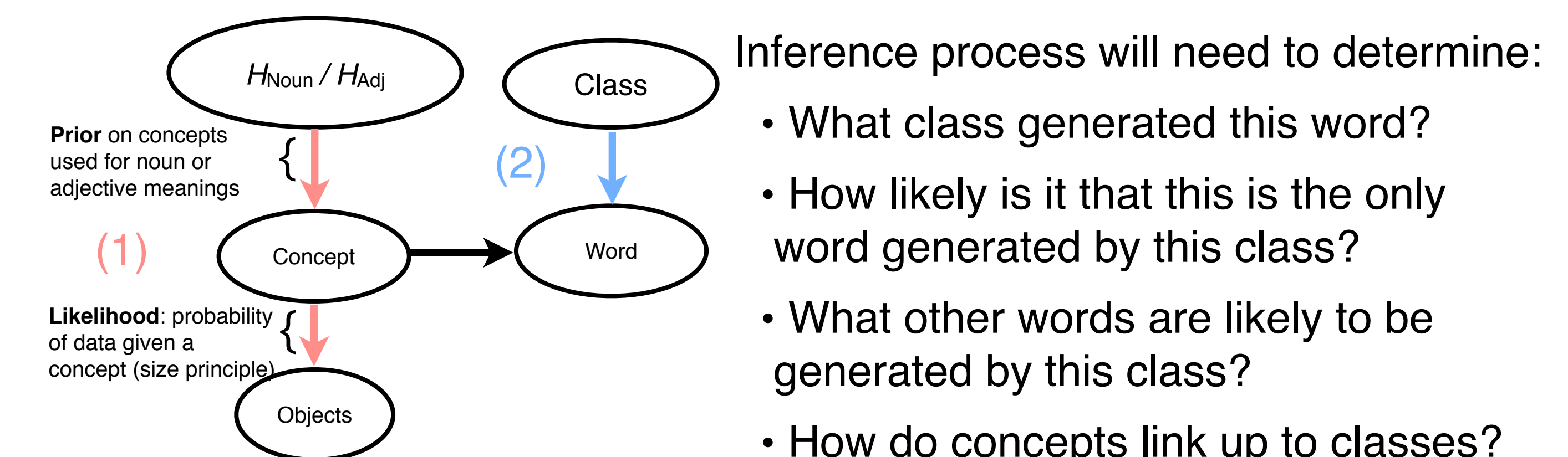
The learner is biased to use different dimensions for nouns and adjective concepts (kind for nouns and property for adjectives) due to the prior probabilities associated with each grammatical category

Adjectives (Exp. 2)

This model would be similar to that proposed for in Exp. 1, except the learner is forced to consider adjective hypotheses that include both kind and property dimensions

Noun Classes

The behavior we see here likely reflects the learner acting on hypotheses determined by 1 of 2 components in the inference process: (1) determine which concept the *one* denotes (2) determine which/how many other concepts are denoted by words in the same class



Inference process will need to determine:

- What class generated this word?
- How likely is it that this is the only word generated by this class?
- What other words are likely to be generated by this class?
- How do concepts link up to classes?