Psych 215L:
Language Acquisition

Lecture 3
Some Child Language Research Methods
&
Some Current Controversies

Statistical Learning, Inductive Bias, & Bayesian Inference in Language Acquisition Research

Statistical learning long considered part of acquisition process (Chomsky 1955, Hayes & Clark 1970, Wolff 1977, Pinker 1984, Goodsitt, Morgan, & Kuhl 1993, among others), but traditionally viewed as playing secondary role rather than primary one.

Why? Children were not believed to be capable of tracking statistical information in language input to the extent that they would need to for learning linguistic knowledge (Chomsky 1981, Fodor 1983, Bickerton 1984, Gleitman and Newport (1995), among others).

Statistical Learning, Inductive Bias, & Bayesian Inference in Language Acquisition Research

Saffran, Aslin, & Newport (1996): groundbreaking study showing experimental support for infant ability to track statistical probability between syllables when trying to segment words from fluent speech.

Saffran et al. proposed that some aspects of acquisition were "best characterized as resulting from innately biased statistical learning mechanisms rather than innate knowledge".

Statistical Learning, Inductive Bias, & Bayesian Inference in Language Acquisition Research

Question 1: What kinds of statistical patterns are human language learners sensitive to?

Thiessen & Saffran (2003): 7-month-olds prefer syllable transitional probability cues over language-specific stress cues when segmenting words, while 9-month-olds show the reverse preference.

Graf Estes, Evans, Alibali, & Saffran (2007): word-like units that are segmented using transitional probability are viewed by 17-month-olds as better candidates for labels of objects.

Thompson & Newport (2007): adults can use transitional probability between grammatical categories to identify word sequences that are in the same phrase, a precursor to more complex syntactic knowledge.

Statistical Learning, Inductive Bias, & Bayesian Inference in Language Acquisition Research

Question 1: What kinds of statistical patterns are human language learners sensitive to?

Other statistics involving relationships of adjacent units: backward transitional probability (Perruchet & Desaulty 2008, Pelucchi, Hay, & Saffran 2009b) and mutual information (Swingley 2005).

Non-adjacent dependencies:

Newport & Aslin (2004): non-adjacent statistical dependencies between consonants and between vowels, but not between entire syllables

Mintz (2002, 2003, 2006): frequent frames used to categorize words. *(ex: the___one is a frame that could occur with big, other, pretty, etc.).*

Statistical Learning, Inductive Bias, & Bayesian Inference in Language Acquisition Research

Question 1: What kinds of statistical patterns are human language learners sensitive to?

More sophisticated statistics/inferences:

Yu & Smith (2007) and Smith & Yu (2008): Both adults and 12- to 14-month-old infants can track probabilities of word-meaning associations across multiple trials where any specific word within a given trial was ambiguous as to its meaning.

Xu & Tenenbaum (2007): investigated how humans learn the appropriate set of referents for basic (*cat*), subordinate *(tabby)*, and superordinate *(animal)* words. Both adults and children between the ages of 3 and 5 are capable of integrating the likelihood of an event occurring into their internal models of word-meaning mapping in a way easily predicted by standard Bayesian inference techniques.

Statistical Learning, Inductive Bias, & Bayesian Inference in Language Acquisition Research

Question 2: To what extent are these statistical learning abilities specific to the domain of language, or even to humans?

Not specific to language:

Saffran et al. (1999): both infants and adults can segment non-linguistic auditory sequences (musical tones) based on the same kind of transitional probability cues that were used in the original syllable-based studies. Similar results have been obtained in the visual domain using both temporally ordered sequences of stimuli (Kirkham et al., 2002) and spatially organized visual "scenes" (Fiser and Aslin, 2002).

Statistical Learning, Inductive Bias, & Bayesian Inference in Language Acquisition Research

Question 2: To what extent are these statistical learning abilities specific to the domain of language, or even to humans?

Not specific to humans:

Hauser et al. (2001): cotton-top tamarins can segment the same kind of artificial speech stimuli used in the original Saffran et al. (1996) segmentation experiments as well as human infants.

Saffran et al. (2008): tamarins could also learn some simple grammatical structures based on statistical information, but were unable to learn patterns as complex as those learned by infants.

Statistical Learning, Inductive Bias, & Bayesian Inference in Language Acquisition Research

Question 3: What kinds of knowledge can be learned from the statistical information available?

Something more easily investigated through computational modeling studies rather than traditional experimental techniques.

Statistical Learning, Inductive Bias, & Bayesian Inference in Language Acquisition Research

The Bayesian approach

- offers a concrete way to examine what knowledge is required for acquisition, and whether that required knowledge is domain-specific or domain-general, without committing to either view *a priori* .

- has led to the investigation of a new set of questions that previous approaches have not considered: whether human language learners can be viewed as being optimal statistical learners (i.e., making optimal use of the statistical information in the data), and in what situations.

- can potentially address the question of why they make the generalizations they do, i.e., because these generalizations are statistically optimal given the available data and any learning biases, innate or otherwise.

Statistical Learning, Inductive Bias, & Bayesian Inference in Language Acquisition Research

The Bayesian approach

- Makes the space of hypotheses considered by the language learner explicit (doesn't matter whether they are based on domain-specific or domain-general cognitive constraints)

- Encodes the learner's biases by assigning an explicit probability distribution over these hypotheses.

- Can operate over the kinds of highly structured representations that many linguists believe are correct (e.g., Regier & Gahl 2004, Perfors, Tenenbaum , & Regier 2006, Foraker et al. 2009, Pearl & Lidz 2009, Perfors et al. to appear).

Statistical Learning, Inductive Bias, & Bayesian Inference in Language Acquisition Research

The Bayesian approach

likelihood of observed data   prior belief in hypothesis

P(hypothesis | data) = P(data | hypothesis) * P(hypothesis)

posterior likelihood of hypothesis   P(data)

likelihood of data period, no matter what hypothesis

"The product of priors and likelihoods often has an intuitive interpretation in terms of balancing between a general sense of plausibility based on background knowledge and the data-driven sense of a "suspicious coincidence." In other words, it captures the tradeoff between the complexity of an explanation and how well it fits the observed data." – Perfors et al. 2010, Bayesian tutorial

Statistical Learning, Inductive Bias, & Bayesian Inference in Language Acquisition Research

The Bayesian approach

Generative framework: observed data are assumed to be generated by some underlying process or mechanism explaining why the data occurs in the patterns it does.
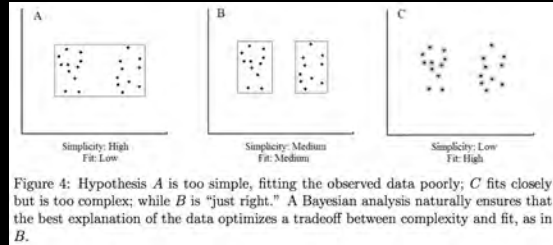Ex: words in a language may be generated by a grammar

Bayesian learner evaluates different hypotheses about the underlying nature of the generative process, and makes predictions based on the most likely ones.

Probabilistic model = a specification of the generative processes at work, identifying the steps (and associated probabilities) involved in generating data.

Statistical Learning, Inductive Bias, & Bayesian Inference in Language Acquisition Research

The Bayesian approach

Figure 4: Hypothesis *A* is too simple, fitting the observed data poorly; *C* fits closely but is too complex; while *B* is "just right." A Bayesian analysis naturally ensures that the best explanation of the data optimizes a tradeoff between complexity and fit, as in *B*.

From Perfors et al. 2010, Bayesian Tutorial

Statistical Learning, Inductive Bias, & Bayesian Inference in Language Acquisition Research
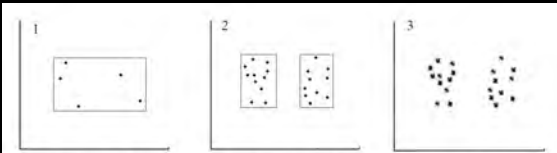
The Bayesian approach

Figure 5: Role of data set size. Three datasets with increasing numbers of datapoints and their corresponding best hypotheses. For data set 1, there are so few datapoints that the simplicity of the hypothesis is the primary consideration; by dat aset 3, the preferred hypothesis is one that fits the clustered datapoints quite tightly.

From Perfors et al. 2010, Bayesian Tutorial

Statistical Learning, Inductive Bias, & Bayesian Inference in Language Acquisition Research

The Bayesian approach

Usual three steps of a Bayesian model:

1) Define hypothesis space – which hypotheses are under consideration?

2) Define prior distribution over hypotheses – which are more/less likely?

3) Define likelihood update – how does data affect learner's belief?

Statistical Learning, Inductive Bias, & Bayesian Inference in Language Acquisition Research

The Bayesian approach

Hypothesis space can contain multiple levels of representation – shows power of bootstrapping (using preliminary or uncertain information in one part of the grammar to help constrain learning in another part of the grammar, and vice versa)

Goldwater et al. (2006, 2009): two levels of representation -- words and phonemes -- though only one of these (words) is unobserved in the input and must be learned.

Johnson (2008): learning both syllable structure and words from unsegmented phonemic input improved word segmentation in a Bayesian model similar to that of Goldwater et al.

Feldman et al. (2009): simultaneously learning phonetic categories and the lexical items containing those categories led to more successful categorization than learning phonetic categories alone.

Statistical Learning, Inductive Bias, & Bayesian Inference in Language Acquisition Research

The Bayesian approach

A note on hierarchical Bayesian models: Allow generalizations at multiple levels.
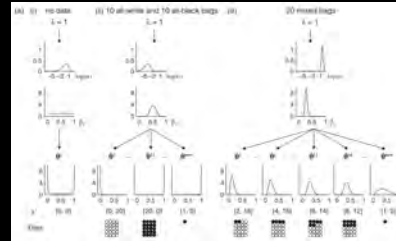
Learner uses observable data to learn about properties of bags in general (ex: uniform vs. mixed distribution), not just properties of individual bags.

Analogy:
bags = language properties

From Kemp, Perfors, & Tenenbaum (2007)

Statistical Learning, Inductive Bias, & Bayesian Inference in Language Acquisition Research

The Bayesian approach

Note: intended to provide a declarative description of what is being learned, not necessarily how the learning is implemented.

Instead: only assume that the human mind implements some type of algorithm (perhaps a very heuristic one) that is able to approximately identify the posterior distribution over hypotheses.

Some studies looking at how Bayesian inference might be implemented:
- Pearl, Goldwater, and Steyvers (2010): implementing Bayesian inference in constrained learners with limitations on memory and processing
- Shi, Griffiths, Feldman, & Sanborn (to appear): exemplar models may provide a possible mechanism for implementing Bayesian inference, and have identifiable neural correlates.

Statistical Learning, Inductive Bias, & Bayesian Inference in Language Acquisition Research

The Bayesian approach

A main contribution: provide a way to formally evaluate claims about children's hypothesis space.

- Can indicate if certain constraints or restrictions are required in order to learn some aspect of linguistic knowledge (e.g., Regier & Gahl 2004, Perfors, Tenenbaum, & Regier 2006, Foraker et al. 2009, Pearl & Lidz 2009, Perfors et al. to appear).

- If a Bayesian learner looking for the optimal hypothesis given the data cannot converge on the correct hypothesis, this suggests that the current conception of the hypothesis space cannot be correct. Required knowledge may take the form of an additional constraint on the hypothesis space that gives preference to certain hypotheses over others, or eliminates some hypotheses entirely.

Statistical Learning, Inductive Bias, & Bayesian Inference in Language Acquisition Research

The Bayesian approach in many different linguistic domains

- Phonetics & perceptual learning: Feldman, Griffiths, & Morgan 2009

- Word segmentation: Goldwater, Griffiths, & Johnson 2009, Pearl, Goldwater, & Steyvers 2010

- Word-meaning mapping: Xu & Tenenbaum 2007, Frank, Goodman, & Tenenbaum 2009

- Syntax-semantics mapping: Regier & Gahl 2004, Pearl & Lidz 2009, Foraker, Regier, Khetarpal, Perfors, & Tenenbaum 2009

- Syntactic structure: Perfors, Tenenbaum, & Regier 2006, Perfors, Tenenbaum, Gibson, & Regier 2010

Current Controversies: Newport 2010

Numbers and symbols

Formal Linguistics: representations and processes are not inherently statistical, but rather are comprised of symbols and rules (Chomsky 1965, 1981, 1995, Marcus 2001). Separation between competence (~symbolic knowledge) and performance (~probabilistic usage of this knowledge).

Psycholinguistics: representations and processes are inherently probabilistic, and symbolic knowledge is more an abstraction imposed by scientists viewing the problem of cognition from the outside. It doesn't make sense to separate knowledge of language from usage of language.

Current Controversies: Newport 2010

Numbers and symbols

Rules vs. Statistics:
- Different types of mental computation (Marcus 1999, Peña et al. 2002)
- Initial representations are more probabilistic while later representations are more symbolic (Hudson Kam & Newport 2009, Newport 1999)

Open question:
- How do humans maintain these two different kinds of knowledge (what is the end state for acquisition)?

Current Controversies: Newport 2010

Modularity

Domain-specificity:
- Is language different from nonlinguistic cognition? Universal Grammar is a particular kind of domain-specific knowledge argued to be special for language.

Separability:
- Are there distinct and modularized components of linguistic processing (ex: phonology vs. syntax)? If so, are there different representations in each module (that children must learn)?