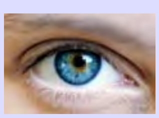
Psych 215L: Language Acquisition

Lecture 2
The Mechanism of Acquisition
and
Some Child Language Research Methods

Levels of Representation Marr (1982)

Describing vs. Explaining

"…it gradually became clear that something important was missing that was not present in either of the disciplines of neurophysiology or psychophysics. The key observation is that neurophysiology and psychophysics have as their business to *describe* the behavior of cells or of subjects but not to *explain* such behavior….What are the problems in doing it that need explaining, and what level of description should such explanations be sought?" - Marr (1982)



On Explaining (Marr 1982)

"…[need] a clear understanding of what is to be computed, how it is to be done, the physical assumptions on which the method is based, and some kind of analysis of the algorithms that are capable of carrying it out."

"This was what was missing - the analysis of the problem as an information-processing task. Such analysis does not usurp an understanding at the other levels - of neurons or of computer programs - but it is a necessary complement to them, since without it there can be no real understanding of the function of all those neurons."

On Explaining (Marr 1982)

"But the important point is that if the notion of different types of understanding is taken very seriously, it allows the study of the information-processing basis of perception to be made *rigorous*. It becomes possible, by separating explanations into different levels, to make explicit statements about what is being computed and why and to construct theories stating that what is being computed is optimal in some sense or is guaranteed to function correctly. The ad hoc element is removed…"

On Explaining (Marr 1982)

"But the important point is that if the notion of different types of understanding is taken very seriously, it allows the study of the information-processing basis of perception to be made *rigorous*. It becomes possible, by separating explanations into different levels, to make explicit statements about what is being computed and why and to construct theories stating that what is being computed is optimal in some sense or is guaranteed to function correctly. The ad hoc element is removed…"

Our goal: Substitute "language acquisition" for "perception".

The three levels

Computational
What is the goal of the computation? What is the logic of the strategy by which it can be carried out?

Algorithmic
How can this computational theory be implemented in a procedure? What is the representation for the input and output, and what is the algorithm for the transformation?

Implementational
How can the representation and algorithm be realized physically?

The three levels: An example with the cash register

Computational
What does this device do?
Arithmetic (ex: addition).
Addition: Mapping a pair of numbers to another number.

(3,4) ⟶ 7 (often written (3+4=7))

True no matter how numbers are represented: this is what is being computed

Properties:
(3+4) = (4+3) [commutative]
(3+4)+5 = 3+(4+5) [associative]
(3+0) = 3 [identity element]
(3+ -3) = 0 [inverse element]

The three levels: An example with the cash register

Computational
What does this device do?
Arithmetic (ex: addition).
Addition: Mapping a pair of numbers to another number.

Algorithmic
What is the input, output, and method of transformation?
Input: arabic numerals (0,1,2,3,4…)
Output: arabic numerals (0,1,2,3,4…)
Method of transformation: rules of addition, where least significant digits are added first and sums over 9 have their next digit carried over to the next column

```
  99
+  5
```

The three levels: An example with the cash register

Computational
What does this device do?
Arithmetic (ex: addition).
Addition: Mapping a pair of numbers to another number.

Algorithmic
What is the input, output, and method of transformation?
Input: arabic numerals (0,1,2,3,4…)
Output: arabic numerals (0,1,2,3,4…)
Method of transformation: rules of addition, where least significant digits are added first and sums over 9 have their next digit carried over to the next column

```
  99
+  5
  14
```

The three levels: An example with the cash register

Computational
What does this device do?
Arithmetic (ex: addition).
Addition: Mapping a pair of numbers to another number.

Algorithmic
What is the input, output, and method of transformation?
Input: arabic numerals (0,1,2,3,4…)
Output: arabic numerals (0,1,2,3,4…)
Method of transformation: rules of addition, where least significant digits are added first and sums over 9 have their next digit carried over to the next column

```
  1
  99
+  5
   4
```

The three levels: An example with the cash register

Computational
What does this device do?
Arithmetic (ex: addition).
Addition: Mapping a pair of numbers to another number.

Algorithmic
What is the input, output, and method of transformation?
Input: arabic numerals (0,1,2,3,4…)
Output: arabic numerals (0,1,2,3,4…)
Method of transformation: rules of addition, where least significant digits are added first and sums over 9 have their next digit carried over to the next column

```
  1
  99
+  5
 104
```

The three levels: An example with the cash register

Computational
What does this device do?
Arithmetic (ex: addition).
Addition: Mapping a pair of numbers to another number.

Algorithmic
What is the input, output, and method of transformation?
Input: arabic numerals (0,1,2,3,4…)
Output: arabic numerals (0,1,2,3,4…)
Method of transformation: rules of addition

Implementational
How can the representation and algorithm be realized physically?
A series of electrical and mechanical components inside the cash register.

The three levels

Marr (1982)

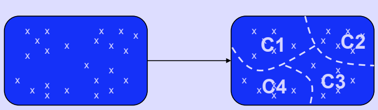
"Although algorithms and mechanisms are empirically more accessible, it is the top level, the level of computational theory, which is critically important from an information-processing point of view. The reason for this is that the nature of the computations that underlie perception depends more upon the computational problems that have to be solved than upon the particular hardware in which their solutions are implemented. To phrase the matter another way, an algorithm is likely to be understood more readily by understanding the nature of the problem being solved than by examining the mechanism (and the hardware) in which it is embodied."

Mapping the Framework: Algorithmic Theory of Language Learning

Goal: Understanding the "how" of language learning

First, we need a computational-level description of the learning problem.

Computational Problem: Divide sounds into contrastive categories
(Speech perception, phoneme identification)



Mapping the Framework: Algorithmic Theory of Language Learning

Goal: Understanding the "how" of language learning

First, we need a computational-level description of the learning problem.

Computational Problem: Divide spoken speech into words
(Word segmentation)

húwzəfɪéjdəvðəbígbǽdwʌ́lf

↓

húwz əfɪéjd əv ðə bíg bǽd wʌ́lf
who's afraid of the big bad wolf

Mapping the Framework: Algorithmic Theory of Language Learning

Goal: Understanding the "how" of language learning

First, we need a computational-level description of the learning problem.

Computational Problem: Identify the concept a word is associated with (Word-meaning mapping)

"I love my daxes."


Dax = that specific toy, teddy bear, stuffed animal, toy, object, …?

Mapping the Framework: Algorithmic Theory of Language Learning

Goal: Understanding the "how" of language learning

First, we need a computational-level description of the learning problem.

Computational Problem: Identify word classes that behave similarly (Grammatical categorization)

"This is a DAX."

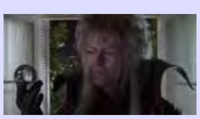
DAX = noun

Mapping the Framework: Algorithmic Theory of Language Learning

Goal: Understanding the "how" of language learning

First, we need a computational-level description of the learning problem.

Computational Problem: Identify the rules of word order for sentences. (Syntax: grammatical rules of the language)



Jareth juggles crystals
Subject Verb Object

Kannada
Subject t_{Object} Verb Object

German
Subject Verb $t_{Subject}$ Object t_{Verb}

English
Subject Verb Object

Mapping the Framework: Algorithmic Theory of Language Learning

Goal: Understanding the "how" of language learning

Second, we need to be able to identify the algorithmic-level description:

- Input = sounds, syllables, words, phrases, …
- Output = sound categories, words, grammatical categories, sentences, …
- Method = statistical learning, algebraic learning, prior knowledge about how human languages work, …

Framework for language learning (algorithmic-level)

What are the hypotheses available (for generating the output from the input)?
Ex: general word order patterns

Input: words (adjective and noun)
Output: ordered pair

Adjective before noun (ex: English)
red apple

Noun before adjective (ex: Spanish)
manzana roja
apple red

Framework for language learning (algorithmic-level)

What are the hypotheses available (for generating the output from the input)?
Ex: general word order patterns

What data are available, and should the learner use all of them?
Ex: exceptions to general word order patterns

Ignore special use of adjective before noun in Spanish
Special use: If the adjective is naturally associated with the noun:
la blanca nieve
the white snow

Why not usual order? Snow is naturally white.

Framework for language learning (algorithmic-level)

What are the hypotheses available (for generating the output from the input)?
Ex: general word order patterns

What data are available, and should the learner use all of them?
Ex: exceptions to general word order patterns

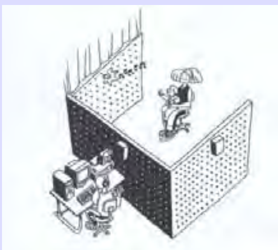
How will the learner update beliefs in the competing hypotheses?
Ex: shifting belief in what the regular word order of adjectives and nouns should be

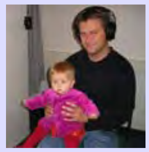
This usually will involve some kind of probabilistic updating function.

Experimental Methods: What, When, and Where

A useful indirect measurement

Head Turn Preference Procedure

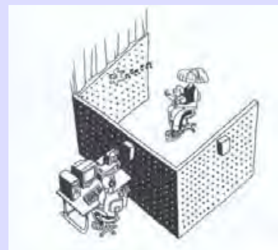


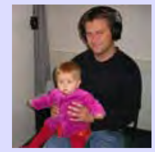


Infant sits on caretaker's lap. The wall in front of the infant has a green light mounted in the center of it. The walls on the sides of the infant have red lights mounted in the center of them, and there are speakers hidden behind the red lights.

A useful indirect measurement

Head Turn Preference Procedure

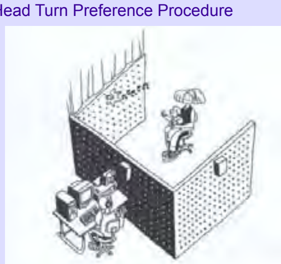


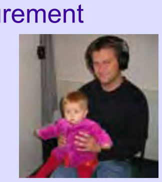


Sounds are played from the two speakers mounted at eye-level to the left and right of the infant. The sounds start when the infant looks towards the blinking side light, and end when the infant looks away for more than two seconds.

A useful indirect measurement

Head Turn Preference Procedure





Thus, the infant essentially controls how long he or she hears the sounds. Differential preference for one type of sound over the other is used as evidence that infants can detect a difference between the types of sounds.

Note on infant attention:

Familiarity vs. Novelty Effects

For procedures that involve measuring where children prefer to look (such as head turn preference), sometimes children seem to have a "familiarity preference" where they prefer to look at something similar to what they habituated to. Other times, children seem to have a "novelty" preference where they prefer to look at something different to what they habituated to.

Kidd, Piantadosi, & Aslin (2010, 2012) provide some evidence that this may have to do with the informational content of the test stimulus. There may be a "Goldilocks" effect where children prefer to look at stimuli that are neither to boring nor too surprising, but are instead "just right" for learning, given the child's current knowledge state.

Computational Methods: How

Computational Methods

Why use computational modeling?

"Given a model of some aspect of language acquisition, implementing it as a computational system and evaluating it on naturally occurring corpora has a number of compelling advantages. First of all by implementing the system, we can be sure that the algorithm is fully specified, and the acquisition model does not resort to hand-waving at crucial points. Secondly, by evaluating it on real linguistic data, we can see whether naturally occurring distributions of examples in corpora provide sufficient information to support the studied claims across a divergent range of acquisition theories. Thirdly, study of the system can identify the mechanisms that cause changes in the algorithm's hypotheses during the course of acquisition. Finally, the computational resources required of the model can be concretely assessed and (not so concretely) compared against the resources that might be available to a human language learner." - Clark & Sakas 2011

Computational Methods

Pearl 2010

Control over the entire learning mechanism:

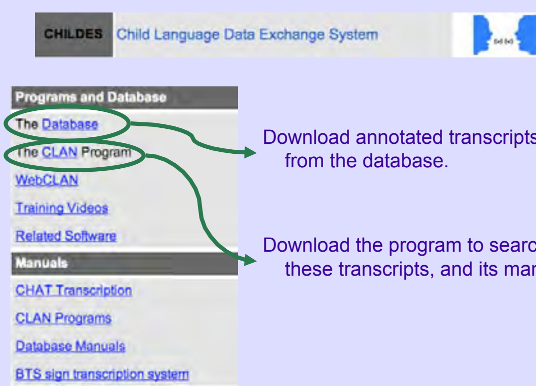
- what hypotheses the (digital) child considers
- what data the child learns from
- how the child updates beliefs in different hypotheses

Ground with empirical data available

- want to make this as realistic as possible (ex: use actual data distributions, cognitively plausible update procedures)
- a good source of empirical data: CHILDES database

http://childes.psy.cmu.edu/

CHILDES Child Language Data Exchange System



Download annotated transcripts from the database.

Download the program to search these transcripts, and its manual.

Back to modeling

Pearl 2010
Gauges of modeling success & contributions to science

Formal sufficiency: does the model learn what it's supposed to learn when it's supposed to learn it from the data it's supposed to learn it from? (also noted as important by Frank 2012; additionally Frank (2012) asks does it make the same mistakes that children do? He calls this fidelity.)

Developmental compatibility: Does it learn in a psychologically plausible way? Is this something children could feasibly do?

Explanatory power: what's the crucial part of the model that makes it work? How does this impact the larger language acquisition story?

Back to modeling

Frank 2012
Additional quality of successful models: Efficient representation

Efficient representation, part 1: "…representations within these models should be efficient compressions of input data at the desired level of analysis…"

Efficient representation, part 2: "…models should include some bias towards parsimony in the representations they learn…a parsimony bias is in its essence, the imposition of some cost on learning such that if one thing is learned, another will not be…"

"…models that [frame] the problem as learning a parsimonious set of explanatory regularities like words, morphemes, categories, or rules—expressive units that [allow] for efficient compression—[are] more successful…"

Sample learning models

Phoneme acquisition (Vallabha et al . 2007, Feldman, Griffiths, & Morgan 2009, Feldman et al. 2011, Elsner et al. 2012, Dillon et al. forthcoming): learning contrastive sounds from acoustic data

Word segmentation (Swingley 2005, Gambell & Yang 2006, Goldwater et al. 2009, Johnson & Goldwater 2009, Blanchard et al. 2010, Jones et al. 2010, Pearl et al. 2011, Lignos 2011, Phillips & Pearl 2012): learning to identify words in fluent speech from streams of syllables

Categorization (Mintz 2003, Wang & Mintz 2008, Chemla et al. 2009, Liebbrandt & Powers 2010, Stumper et al. 2011): learning to identify what category a word is (noun, verb) from segmented speech

Sample learning models

Morphology (Rumelhart & McClelland 1986, Yang 2002, Albright & Hayes 2002, Yang 2005, Chan & Lignos 2011, Gagliardi et al. 2012): learning to identify word affixes from segmented speech

Learning the interpretation of referential elements (Regier & Gahl 2004, Foraker et al. 2007, 2009, Pearl & Lidz 2009, Pearl & Mis 2011): learning to identify syntactic category and semantic referent of *one* from segmented speech and referents in the world

Sample learning models

Syntactic acquisition (Yang 2004, Reali & Christiansen 2005, Kam et al. 2008, Pearl & Weinberg 2007, Perfors, Tenenbaum, & Regier 2011, Pearl & Sprouse 2011): learning to identify correct word order (rules) from speech segmented into words

Stress (Pearl 2008, Pearl 2011, Legate & Yang 2011): learning to identify correct stress patterns (and rules behind them) from words with stress contours

General Modeling Process

Pearl 2010

(1) Decide what kind of learner the model represents (ex: normally developing 6-month-old child learning first language)

(2) Decide what data the child learns from (ex: Bernstein corpus from CHILDES) and how the child processes that data (ex: data divided into syllables)

(3) Decide what hypotheses the child has (ex: what the words are) and what information is being tracked in the input (ex: transitional probability between syllables)

(4) Decide how belief in different hypotheses is updated (ex: based on transitional probability minima between syllables)

General Modeling Process

Pearl 2010

(5) Decide what the measure of success is

- precision and recall (ex: finding the right words in a word segmentation task)

- matching an observed performance trajectory (ex: English past tense acquisition often has a U-shaped curve)

- achieving a certain knowledge state by the end of the learning period (ex: knowing there are 4 vowel categories at the end of a phoneme identification task)

- making correct generalizations (ex: preferring a correctly formed sentence over an incorrectly formed one)

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

Pearl & Goldwater forthcoming

"Language acquisition is a problem of induction: the child learner is faced with a set of specific linguistic examples and must infer some abstract linguistic knowledge that allows the child to generalize beyond the observed data, i.e., to both understand and generate new examples. Many different generalizations are logically possible given any particular set of input data, yet different children within a linguistic community end up with the same adult grammars. This fact suggests that children are biased towards making certain kinds of generalizations rather than others."

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

Pearl & Goldwater forthcoming

"In the Bayesian view of learning, inductive bias consists of a combination of hard and soft constraints. Hard constraints make certain grammars impossible for any human to acquire; in the language of Bayesian modeling, these impossible grammars are outside the learner's *hypothesis space*. Grammars inside the hypothesis space are learnable given the right input data, but they may not all be equally easy to learn. Soft constraints, implemented in the form of a probability distribution over the hypothesis space, mean that the learner will be biased towards certain of these grammars more than others. "

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

Pearl & Goldwater forthcoming

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. (See Romberg & Saffran 2010 for a review of infant statistical learning abilities, and Aslin & Newport 2012 for a review of statistical learning abilities in both infants and adults.)

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

Pearl & Goldwater forthcoming

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

Evidence for domain-general probabilistic learning abilities in infants
- Denison et al. forthcoming: 6-month-olds (prob reasoning)
- Roseberry et al. 2011: 7- to 9-month-olds (prob tracking)
- Davis et al. 2011: 10-month-olds (prob matching)

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 (always) 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

- Also, may be different ways to approximate Bayesian inference that are not so resource-intensive. Bonawitz, Denison, Chen, Gopnik, & Griffiths (2011) discuss a simple sequential algorithm called Win-Stay, Lose-Shift that matches human behavior consistent with Bayesian inference.
- Some evidence that infants are sensitive to certain kinds of information that we would expect Bayesian learners to be sensitive to: Gweon et al. (2010) show that 12- to 18-month-old children alter their inferences, based on where the sample is drawn from.

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, Foraker et al. 2009, Pearl & Lidz 2009, Pearl & Mis 2011, Perfors et al. 2011).

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) / P(data)

posterior likelihood of hypothesis

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. 2011, 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

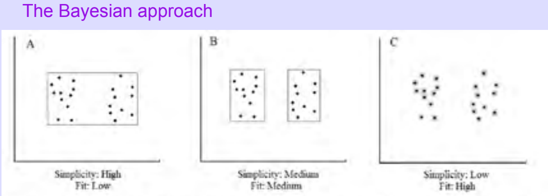


Figure 4: Hypothesis *A* is too simple, fitting the observed data poorly; *C* fits closely but is too complex; while *B* is "just right." A Bayesian analysis naturally ensures that the best explanation of the data optimizes a tradeoff between complexity and fit, as in *B*.

From Perfors et al. 2011, 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.

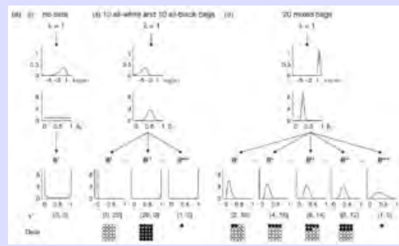
Feldman et al. (2009): simultaneously learning phonetic categories and the lexical items containing those categories led to more successful categorization than learning phonetic categories alone.

Yuan et al. (2011): simultaneously learning individual word meaning and more abstract features involved in word meaning

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

The Bayesian approach

A note on hierarchical Bayesian models: Allow generalizations at multiple levels. (Dewar & Xu (2010): 9-month-olds can do this.)



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.

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

The Bayesian approach

Some studies looking at how Bayesian inference might be implemented:

- Pearl, Goldwater, and Steyvers 2010, 2011, Phillips & Pearl 2012: implementing Bayesian inference in constrained learners with limitations on memory and processing

- Shi, Griffiths, Feldman, & Sanborn 2010: 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 2011, Foraker et al. 2009, Pearl & Lidz 2009, Pearl & Mis 2011, Perfors et al. 2011).

- 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, Feldman et al. 2011, Dillon et al. forthcoming

- Word segmentation: Goldwater, Griffiths, & Johnson 2009, Johnson & Goldwater 2009, Pearl, Goldwater, & Steyvers 2010, 2011

- 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, Pearl & Lidz 2011

- Syntactic structure: Perfors, Tenenbaum, Gibson, & Regier 2010, Perfors, Tenenbaum, & Regier 2011

Extra slides

Experimental Methods

How do we tell what infants know, or use, or are sensitive to?

Researchers use indirect measurement techniques.

High Amplitude Sucking (HAS)

Infants are awake and in a quietly alert state. They are placed in a comfortable reclined chair and offered a sterilized pacifier that is connected to a pressure transducer and a computer via a piece of rubber tubing. Once the infant has begun sucking, the computer measures the infant's average sucking amplitude (strength of the sucks).

Experimental Methods

How do we tell what infants know, or use, or are sensitive to?

Researchers use indirect measurement techniques.

High Amplitude Sucking (HAS)

A sound is presented to the infant every time a strong or "high amplitude" suck occurs. Infants quickly learn that their sucking controls the sounds, and they will suck more strongly and more often to hear sounds they like the most. The sucking rate can also be measured to see if an infant notices when new sounds are played.

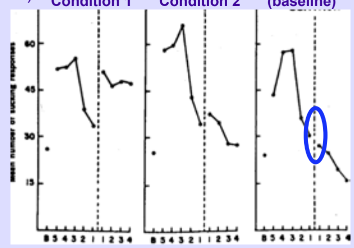
Experimental Methods

How do we tell what infants know, or use, or are sensitive to?

Researchers use indirect measurement techniques.

High Amplitude Sucking (HAS)

Test Condition 1 **Test Condition 2** **Control (baseline)**



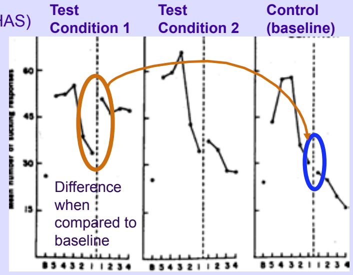
Experimental Methods

How do we tell what infants know, or use, or are sensitive to?

Researchers use indirect measurement techniques.

High Amplitude Sucking (HAS)

Test Condition 1 **Test Condition 2** **Control (baseline)**



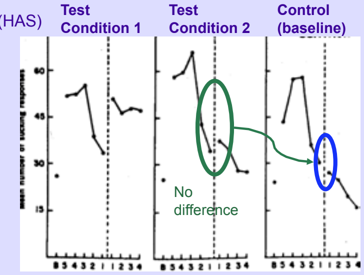
Experimental Methods

How do we tell what infants know, or use, or are sensitive to?

Researchers use indirect measurement techniques.

High Amplitude Sucking (HAS)

Test Condition 1 **Test Condition 2** **Control (baseline)**



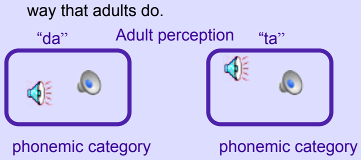
Experimental Methods

How do we tell what infants know, or use, or are sensitive to?

Researchers use indirect measurement techniques.

High Amplitude Sucking (HAS)

Infants have sophisticated discrimination abilities, but they don't abstract sounds into categories the way that adults do.



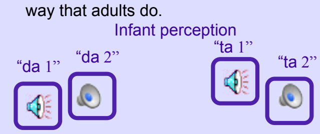
Experimental Methods

How do we tell what infants know, or use, or are sensitive to?

Researchers use indirect measurement techniques.

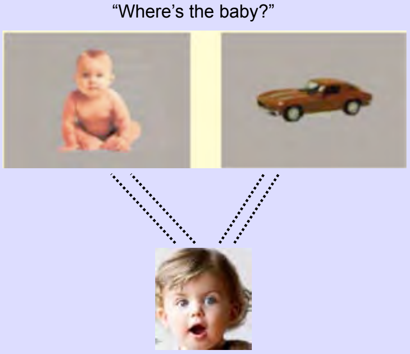
High Amplitude Sucking (HAS)

Infants have sophisticated discrimination abilities, but they don't abstract sounds into categories the way that adults do.



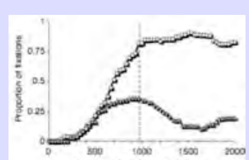
Eyetracking: measures fixations on target picture

"Where's the baby?"



Eyetracking: measures fixations on target picture

"Where's the baby?"



"Where's the **baby**?"

"Where's the **vaby**?"

Looking at children's brains

ERPs: Event-related brain potentials, gauged via electrode caps. The location of ERPs associated with different mental activities is taken as a clue to the area of the brain responsible for those activities.

Good: non-invasive, relatively undemanding on the subject, provide precise timing on brain events

Bad: poor information on exact location of ERP since just monitoring the scalp

Looking at children's brains

Brain-imaging techniques: gauge what part of the brain is active as subjects perform certain tasks

PET scans: Positron emission topography scans
- subjects inhale low-level radioactive gas or injected with glucose tagged with radioactive substance
- experimenters can see which parts of the brain are using more glucose (requiring the most energy)

fMRI scans: functional magnetic resonance imaging
- subjects have to be very still inside MRI machine, which is expensive to operate
- experimenters can see which parts of the brain are getting more blood flow or consuming more oxygen

Looking at children's brains

Brain-imaging techniques: gauge what part of the brain is active as subjects perform certain tasks

MEG: Magnetoencephalography
- subjects have to be very still
- experimenters can see which parts of the brain are active

Looking at children's brains

Brain-imaging techniques: gauge what part of the brain is active as subjects perform certain tasks

Optical Topography: Near-infrared spectroscopy (NIRS)
- transmission of light through the tissues of the brain is affected by hemoglobin concentration changes, which can be detected