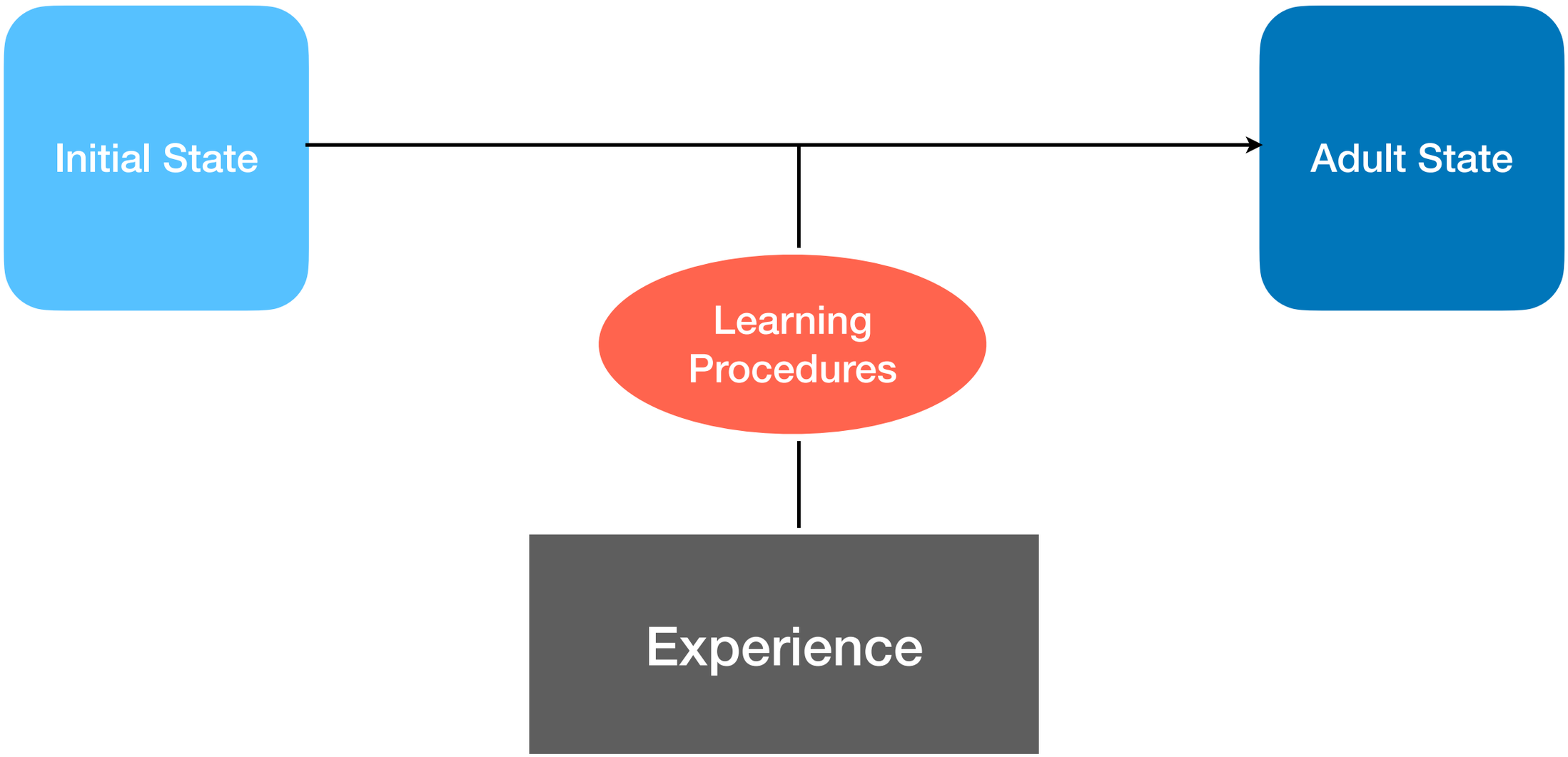
24.904
Language Acquisition

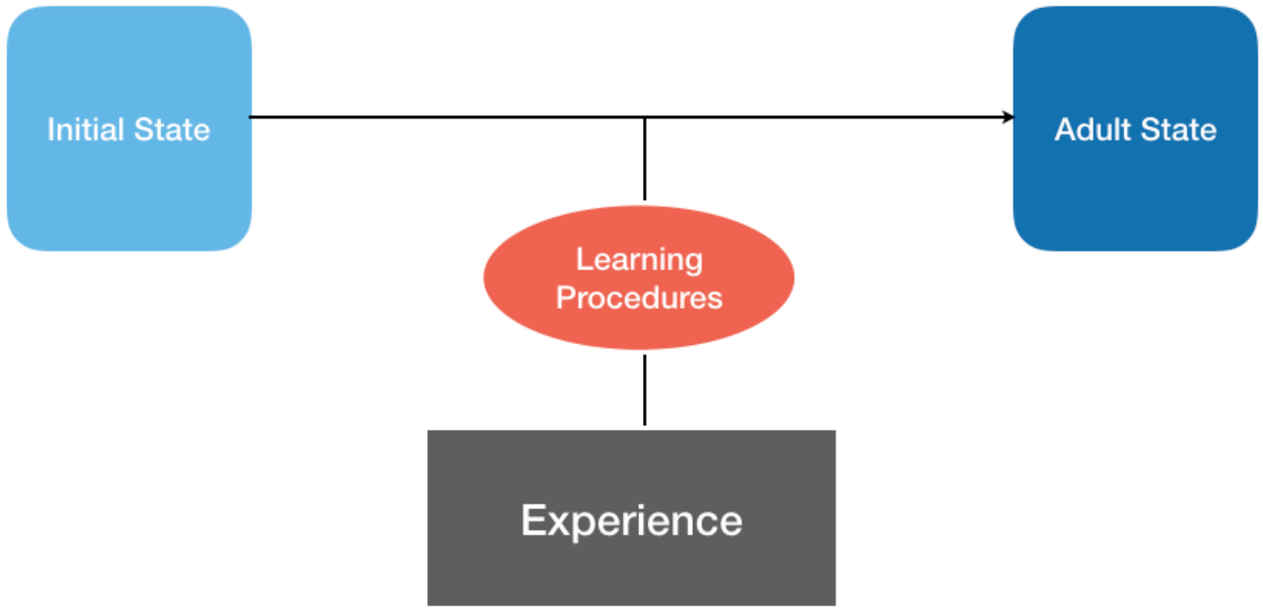
Class 2: Biological Bases of Language

First language acquisition



Division of labor

- What is the role of environment (=experience) ?
- What is the role of biological endowment (=initial state) ?



Navigational knowledge in bees

- Forager bees take an irregular path from hive until a new food source is found

- They then will fly a straight path back to the hive and make a peculiar “dance” inside the hive

- Shortly thereafter, many bees will fly a straight path to the newly discovered food source

Navigational knowledge in bees

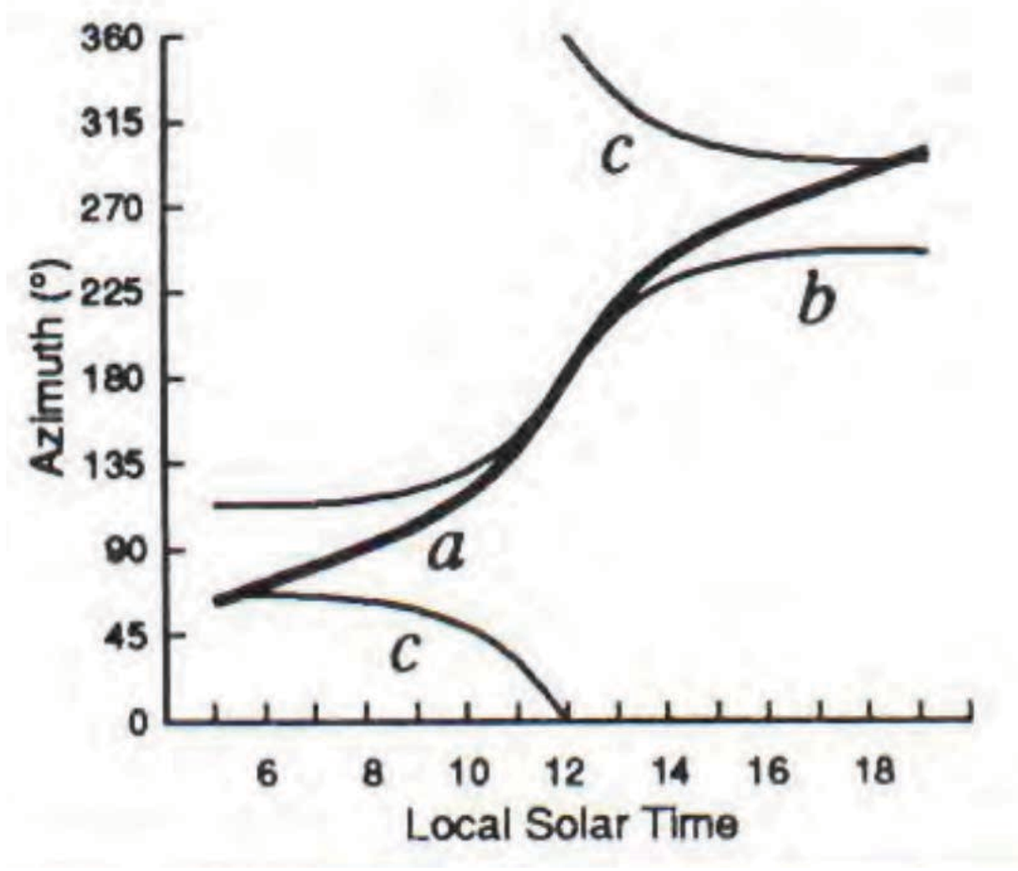
What has to be represented?

- The sun’s movement across the sky varies by location on the earth and season

- The sun’s movement across the sky varies by time of day

- **Local ephemeris function:** The function that describes the sun's position with time at a given latitude

- Given that bees use the sun for navigation and communication, they must have a working sense of the ephemeris function.



Local ephemeris function for East Lansing, MI (D&D96)

How do they acquire this knowledge?

- How do bees form an internal representation of the solar ephemeris from experiencing seeing the sun moving relative to the terrain?

- Not a memorized list of time-linked positions

  - can estimate sun's position for times of day when they have never seen it (Lindauer et al. 1959)

Dyer & Dickinson 1996

- Incubator-reared bees who were allowed to see only a small (~20%) portion of the sun's daily course (within the 4hr period before sunset)

- examined how these afternoon-experienced bees estimated the course of the sun in the morning and middle of the day (by observing their dances).

Dyer & Dickinson 1996

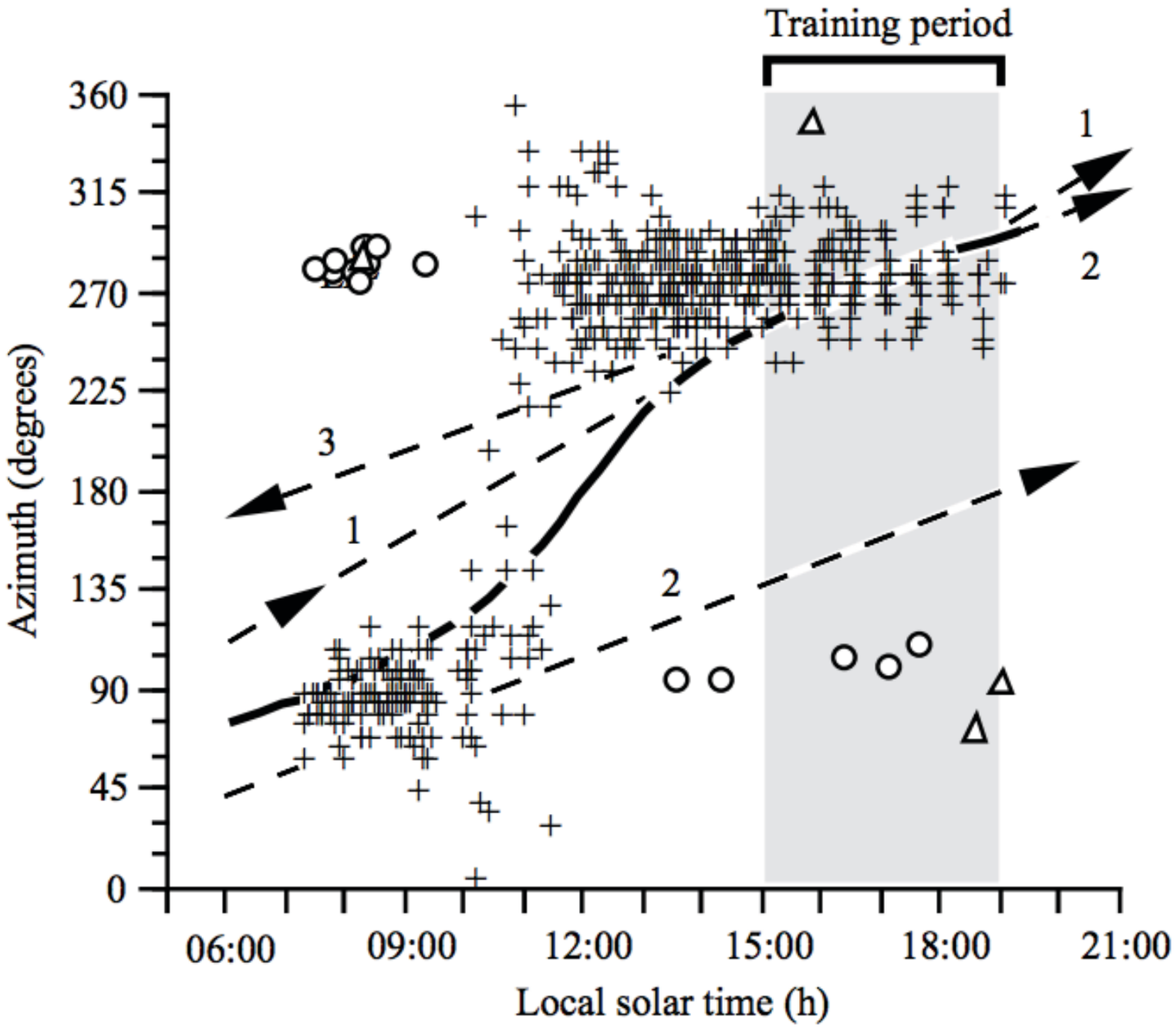
- Three hypotheses:

1. bees might ‘interpolate’ at a linear rate to find the sun’s position between two known positions

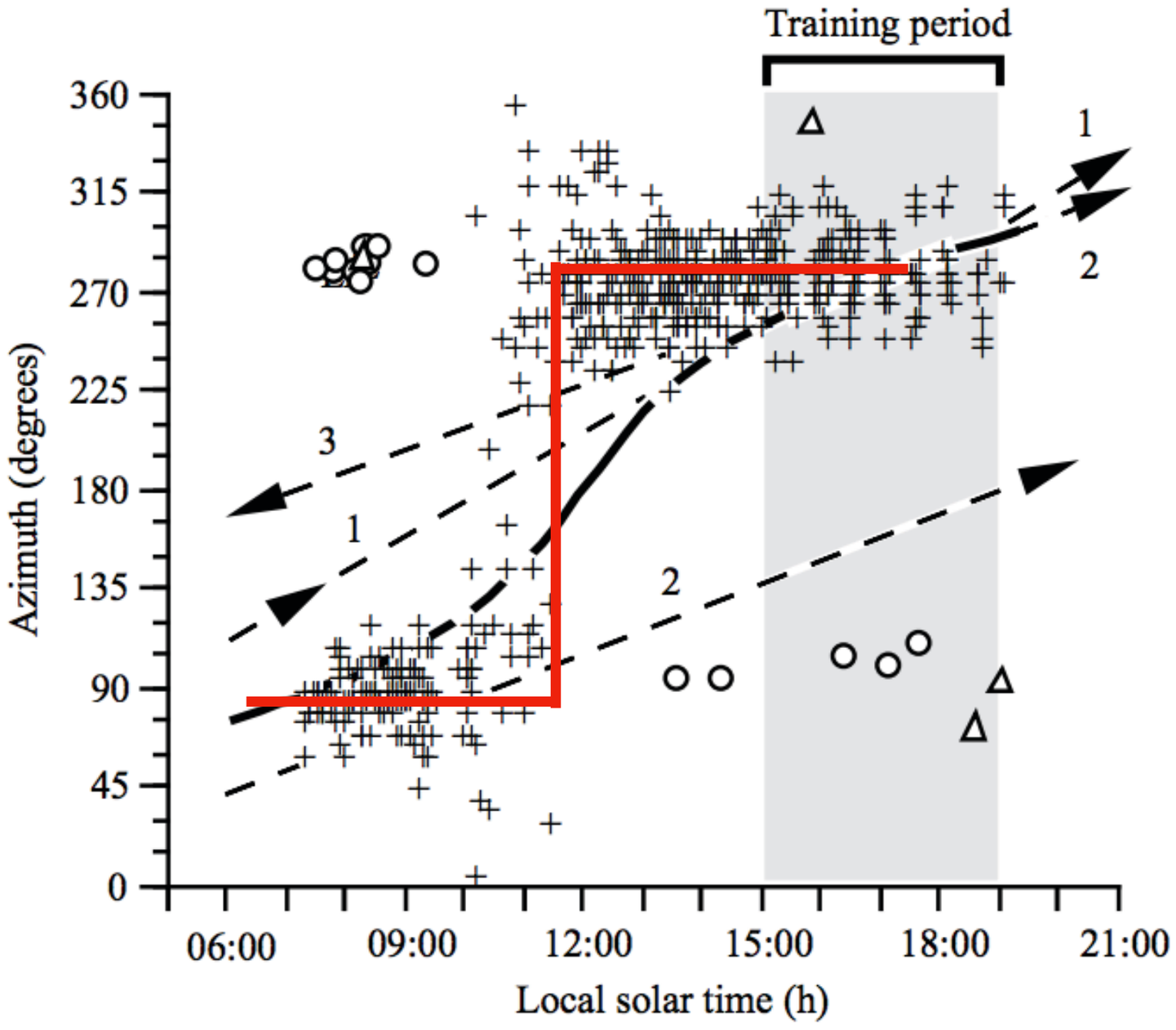
2. extrapolate forward into the morning hours at a linear rate based on the most recently observed rate of movement

3. extrapolate backwards at a linear rate into the morning hours of the sun’s rate as measured at the beginning of the training period on previous days

Dyer and Dickinson 1996



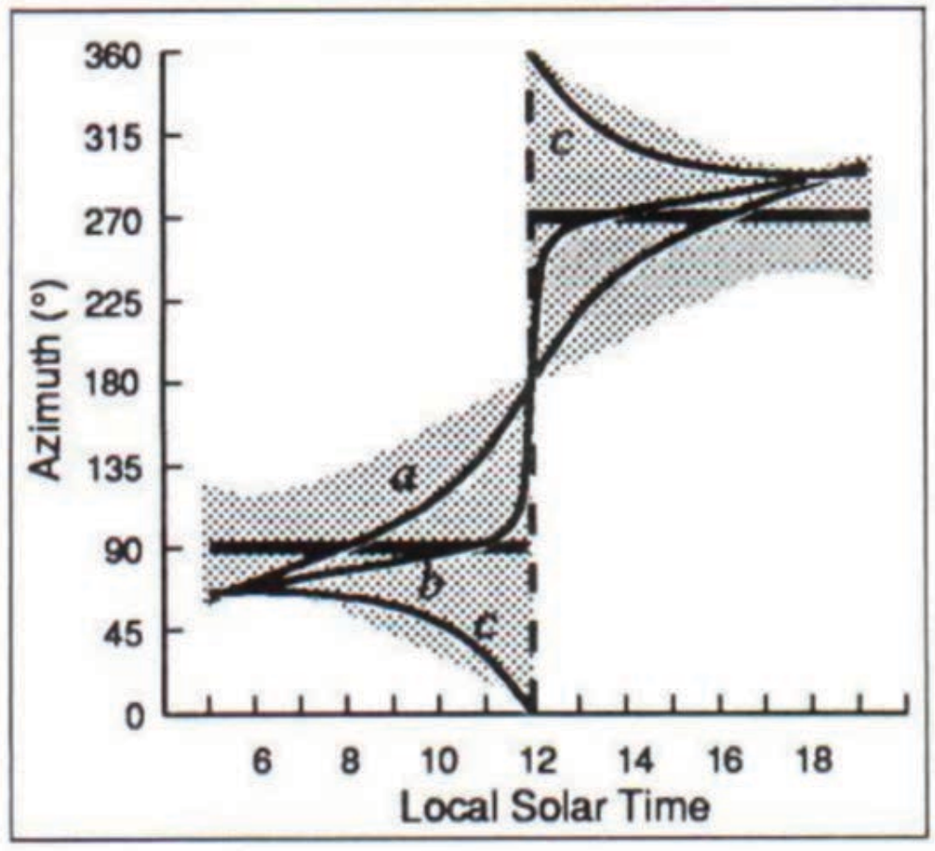
Dyer and Dickinson 1996



Dyer & Dickinson 1996

- In the morning, they used an azimuth ~180˚ from what they had learned from training. In the afternoon, they shifted by ~180˚

- Rather than some constant rate of movement, bees behaved as if they used an (imperfect/incomplete) representation of the sun's actual non-linear pattern of movement

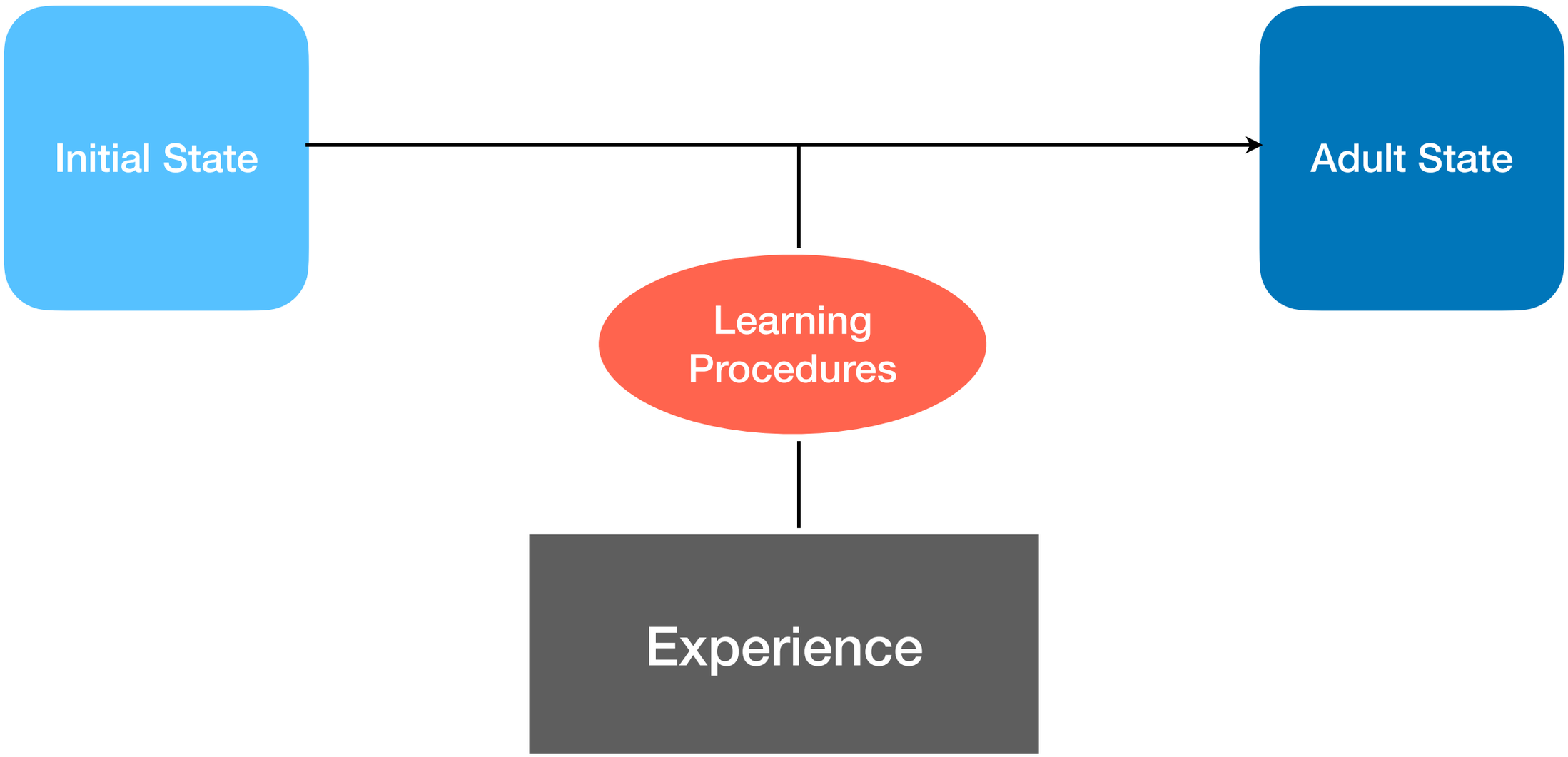


Actual patterns of solar movement on Earth, with shaded area representing all possible solar ephemeris functions

Acquisition of navigational knowledge in bees

- An innate template
  - representation of an approximate ephemeris in the shape of a 180° step function
- Learning as filling in that template
  - to refine, with experience, the shape of the particular relevant local ephemeris

First language acquisition



Characterizing the learning problem

- The output of learning is complex
- The output of learning not always easy to observe
- The input for learning not always easy to observe
- Yet learning is **robust** and **fast**

Output of learning is complex

- A native speaker of English apparently knows:

(2) a. Who do you think {∅/that} your cat likes?
b. Who do you think {∅/*that} likes your cat?

and…

(3) a. The man {∅/that} your cat likes is nice.
b. The man {*∅/that} likes your cat is nice.

and varies in complex ways across speakers/speaker communities

On the surface variation:

- English speakers know an SVO language, Japanese speakers know an SOV language

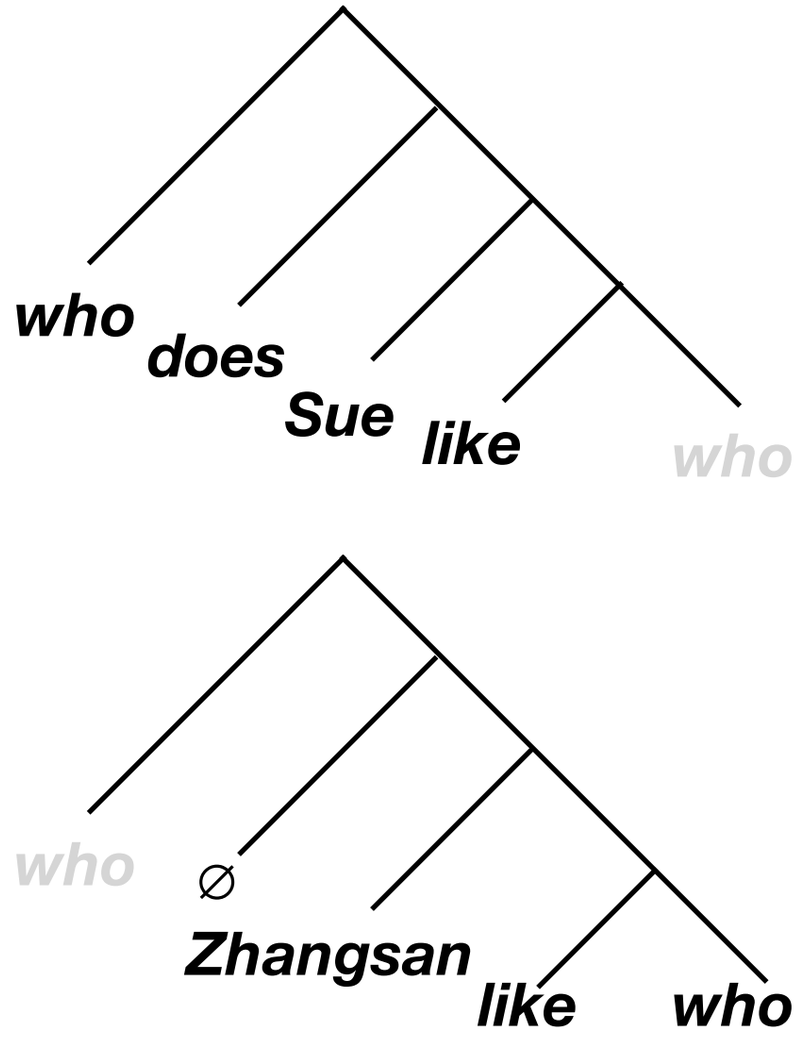
(3) a. John ate a pizza *vs.*
b. John-ga piza-o tabeta

and varies in complex ways across speakers/speaker communities

On the surface variation:

- English *wh*-questions involve wh-fronting, Mandarin does not

(4) a. Who does Sue like? *vs.*
b. Zhangsan xihuan shei?
Z. like who



in ways that are not surface-evident

On-the-surface, but less straightforward, variation:

- Verb placement in English vs. French

(5) a. Marie does **not speak** French *vs.* English
b. Marie ne **parle pas** français French

but also…

(6) a. …to **not speak** French *vs.* English
b. …ne **pas parler** français French

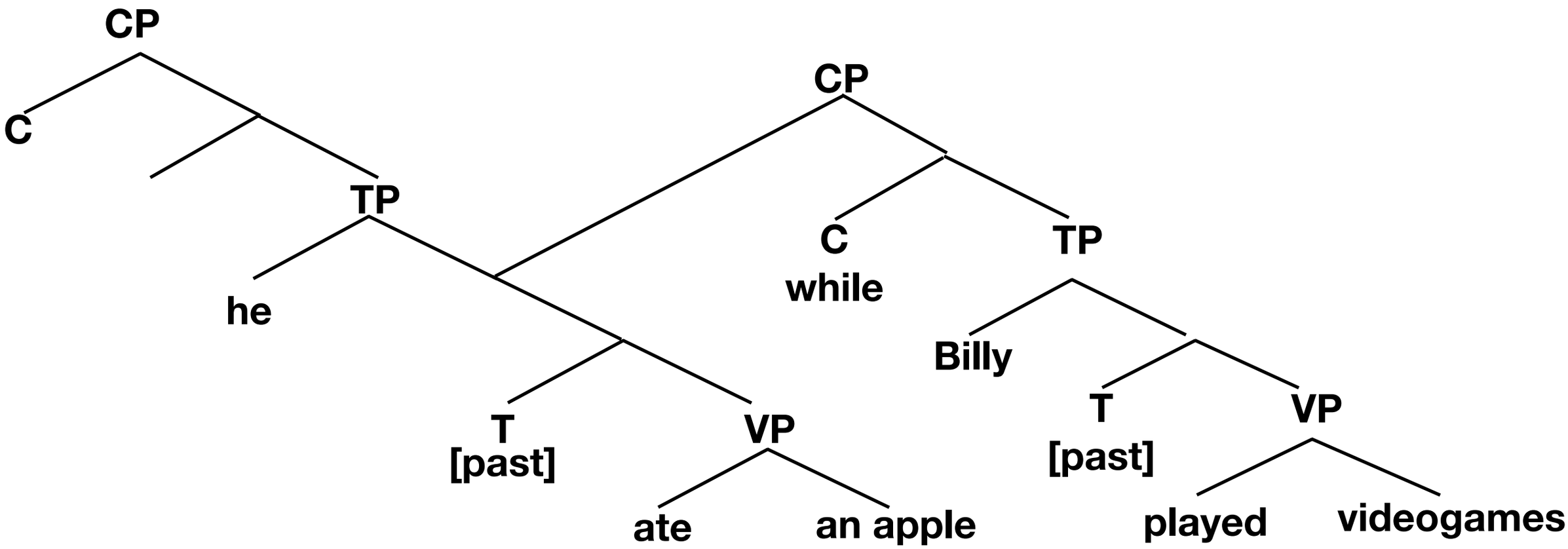
in ways that are not surface-evident

Below-the-surface variation:

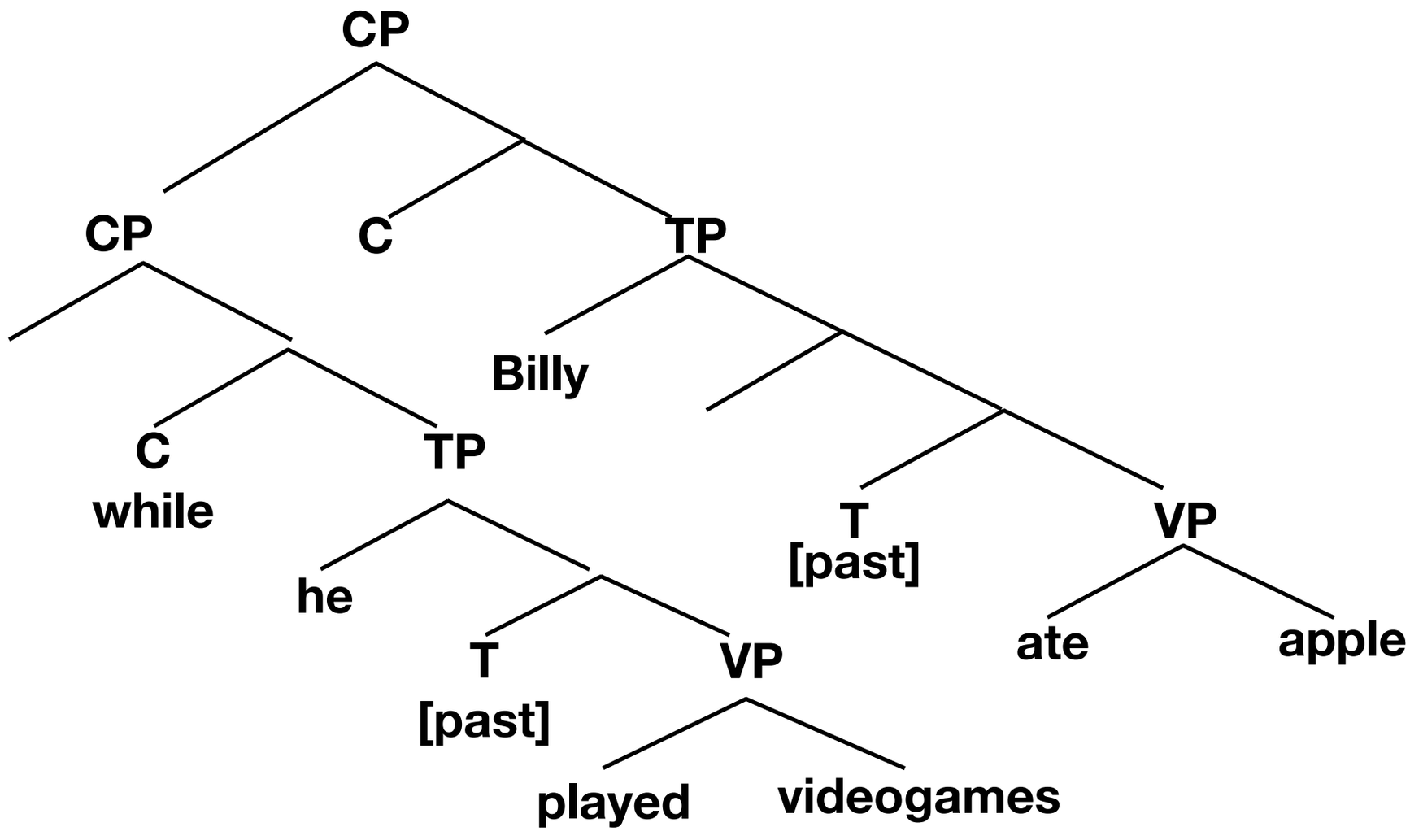
(7) a. Billy$_i$ ate an apple while he$_i$ was playing videogames.
b. He$_i$ ate an apple while Billy$_i$ was playing videogames.

(8) a. While Billy$_i$ was playing videogames, he$_i$ ate an apple.
b. While he$_i$ was playing videogames, Billy$_i$ ate an apple.

he c-commands Billy



he does not c-command Billy



in ways that are not surface-evident

Below-the-surface variation:

(7) a. ✓English, ✓Russian, ✓Malayalam
Billy$_i$ ate an apple while he$_i$ was playing videogames.
b. ✖English, ✖Russian, ✖Malayalam
He$_i$ ate an apple while Billy$_i$ was playing videogames.

(8) a. ✓English, ✓Russian, ✓Malayalam
While Billy$_i$ was playing videogames, he$_i$ ate an apple.
b. ✓English, ✖Russian, ✖Malayalam
While he$_i$ was playing videogames, Billy$_i$ ate an apple.

Yet learning is robust

Uniform across individuals…

- Linguistic milestones achieved in parallel fashion
  - babbling around 6-8 months
  - first word production around 10-12 months
  - spurt of vocabulary growth and productive word combinations in the second year of life

and fast

- Infants as young as 6mos show understanding of several words[1]

- Children's earliest multiword utterances (~1;6yo) reveal knowledge of language-specific syntactic properties
  - Knowledge of head-directionality (OV vs. VO)[2]
  - Knowledge of verb-placement relative to negation[3]

✳ **By ~age 6, child language virtually indistinguishable from adult language**

1 Bergelson 2012, Bergelson & Swingley 2013
2 Bloom 1970, Brown 1973, Clahsen 1985, Wexler 1994, 1998, Sugisaki 2005, 2008
3 Stromswold 1990, Pierce 1992, Deprez & Pierce 1993, Harris & Wexler 1996

Characterizing the learning problem

- The output of learning is complex
- The output of learning not always easy to observe
- The input for learning not always easy to observe
- Yet learning is **robust** and **fast**

The hypothesis

- We are born already knowing something about how language works

- i.e. the learner has **some initial delimitation of a class of possible hypotheses** and **a method of determining what each such hypothesis implies** with respect to each encountered sentence

- One variant (implies domain-specificity):
  - our minds come “pre-loaded” with a kind of template for linguistic rules and representations
  - Language acquisition = filling in this template with experience

The logical argument

- Poverty of the stimulus
  - Our experience underdetermines our knowledge
  - Hence, something experience-independent must be partly responsible for the derived state

The logical argument

- Why does linguistic experience underdetermine linguistic knowledge?

The logical argument

Why does linguistic experience underdetermine linguistic knowledge?

1. The output of acquisition and the input of acquisition are ontologically different kinds of objects.

   - Input/ambient linguistic data are **uttered sentences/phrases**

   - Native speakers are linguistically creative, i.e. they can understand and produced never-before encountered linguistic objects

   - Linguistic creativity requires that what is acquired on the basis of experience is not a set of previously encountered sentences, but a generative procedure (rule system or **grammar**) that can recursively specify the *open ended* list of acceptable sentences/phrases of the speaker's L1

Aside…

- If one denies that what is learned is a grammar (some people do), then one can deny the POS argument in an internally consistent manner

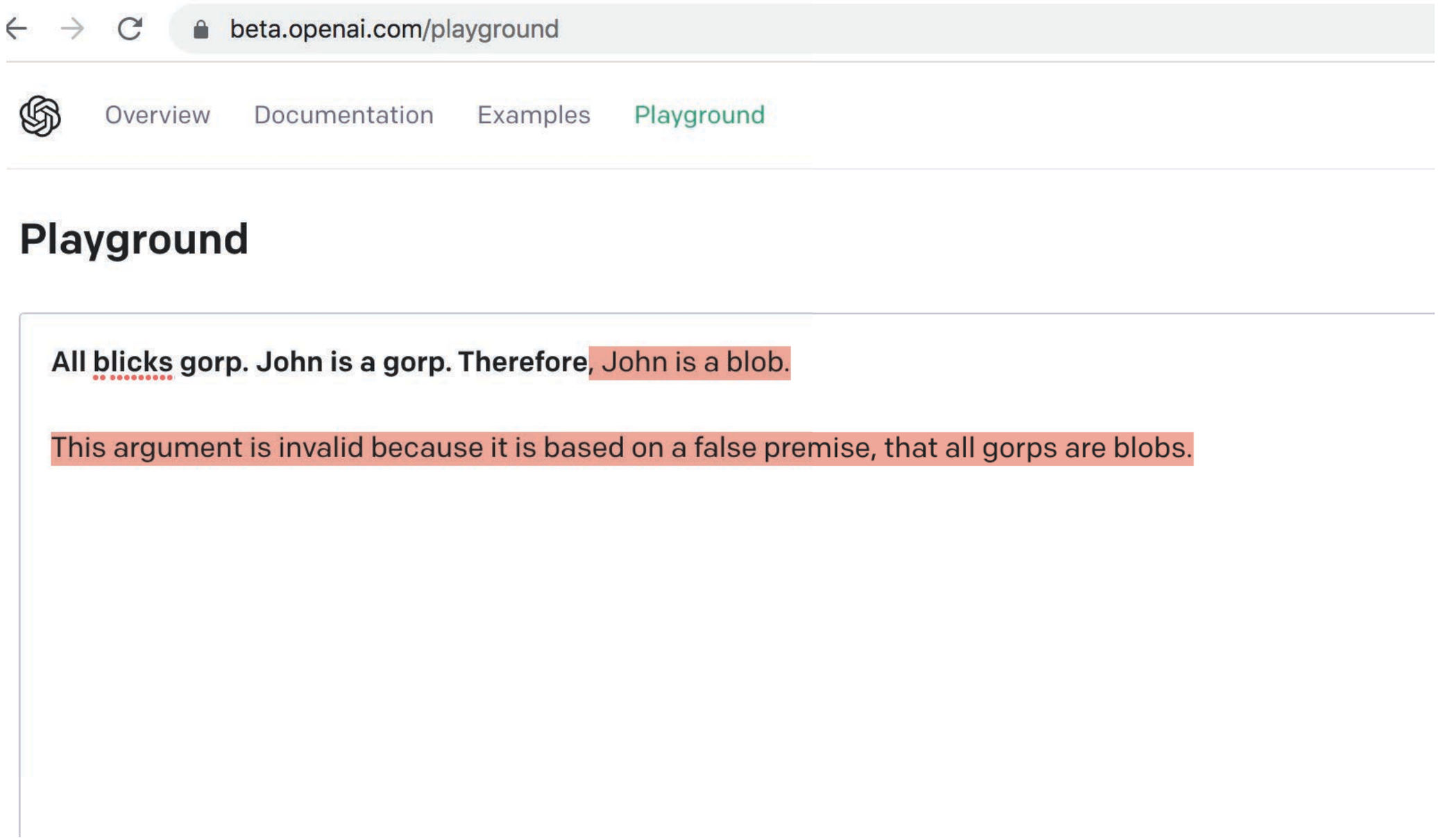
- E.g. we have neural net models that (e.g. GPT-2) are remarkably sophisticated. But is the linguistic knowledge it has the same kind of knowledge that you have?

All blicks gorp. John is a blick. Therefore, John…

https://talktotransformer.com/

Aside…

Getting better…



The logical argument

Why does linguistic experience underdetermine linguistic knowledge?

2. There is a gap between the grammars that are projected and the evidence to choose among plausible grammars

   - For any finite set of data specified extensionally, there are an infinite number of different functions that can generate those data.

   - The data itself does not specify ways of choosing one grammar over the other

   - In fact it doesn't even contain information about the kinds of representations that should be used to build the right sorts of grammars

   - Yet speakers within a speaker community converge on a grammar, suggesting that they are not guessing.

Some conceptual arguments

- **Species-specific:** only humans
- Within the species, **uniform**:
  - all languages are comparably complex
  - all humans can acquire any of the world's languages

Argument from typology

When we look across the world's languages, we find…

Striking similarities:

- Set of phonological features, e.g. [±voice], [±nasal]
- Set of categories (N, V, D, T, C)
- Set of operations (Merge, selection, modification)
- Order of operations (Merge V before T before)

Argument from typology

When we look across the world's languages, we find…

Highly restricted differences:

- A language may front its *wh*-words (English) or leave it in the same place as the answer (Mandarin), but no language systematically pronounces *wh*-words at the end of the sentence.

- Languages may have V2 (German) or not, but no language has V3

  … though caveats about (non-)universality arguments

Argument from the relative *irrelevance* of experience

- Cases where radical alterations in the environment seemingly has no effect on the fundamental character of language acquisition

- Topic of next week.

  - Read: Senghas, Bedny et al.