Simulating Language
15: Bayesian Learning

Simon Kirby

simon@ling.ed.ac.uk

The story so far, and what comes next

- Signalling systems (and languages) can evolve as a result of their transmission
  - We can model this
- The **biases** of learners shapes what evolves
- This potentially allows us to link findings about biases in learning at the individual level to predictions / observations about language at the population level
  - But caution (or better, a model) is required - the acquisition test here was misleading, there is more than one route to optimal systems
- **Next up:** a class of models that allow us to be very clear and very precise about bias

We need a more general model

- Ideally, we'd like to be able to mix up a bunch of simple ingredients and work out what language should look like after cultural evolution has run for some time:
  - BIAS (i.e. what agents are born with)
  - LANGUAGE MODEL (i.e. set of possible languages, set of possible data)
  - POPULATION MODEL (e.g. diffusion chain, closed group etc.)
  - OTHER FEATURES OF CULTURAL TRANSMISSION (e.g. how much data a learner sees, the type of errors that occur, ...)

Towards a more general model of learning bias: a medical quiz

- Your friend coughs. Is this cough caused by:

A. SARS (a deadly flu-like virus spread from cave-dwelling bats)

B. A cold

C. Athlete's foot

- Resolving this question requires you to draw on two probabilities:
  - How likely is it that someone with the illness in question would exhibit that symptom?
  - How common is each illness?

Likelihood of symptoms given illnesses

SARS: coughing is very likely, if you have SARS

A cold: coughing is very likely, if you have a cold

Athlete's foot: coughing is very very unlikely to be caused by athlete's foot

- If all we care about are the likelihood of the symptoms given each illness, we would conclude that your friend either has lung cancer or a cold

Probability of illnesses

SARS: is very rare

A cold: the common cold is very common

Athlete's foot: is very common (let's say)

- If all we care about are the prevalence of each illness, we would conclude that your friend either has a cold or athlete's foot
- But you didn't conclude this: you brought these two quantities together in a smart way. How did you do it?

The Bayesian approach

- What you're trying to figure out is the probability that your friend has a particular illness, given the symptoms they are exhibiting. We call this quantity:

$$P(illness|symptoms)$$

- We are trying to work this out based on two quantities which we know (roughly):
  - The likelihood of exhibiting a particular symptom given that you have a certain illness

$$P(symptoms|illness)$$

  - The prior probability of each illness

$$P(illness)$$

Bayes' rule

- Bayes' rule provides a convenient way of expressing the quantity we want to know in terms of the quantities we already know:

$$P(illness|symptoms) \propto P(symptoms|illness)P(illness)$$

- Or, in full:

$$P(illness|symptoms) = \frac{P(symptoms|illness)P(illness)}{P(symptoms)}$$

Breaking it down

$$P(illness|symptoms) = \frac{P(symptoms|illness)P(illness)}{P(symptoms)}$$

$P(illness|symptoms)$

- The thing we want to know is called the **posterior**

$P(symptoms|illness)$

- The probability of a particular set of symptoms given that you have a specific illness is called the **likelihood**

$P(illness)$

- The probability that you have a particular illness, before I have any evidence from your symptoms, is called the **prior**

$P(symptoms)$

- The term on the bottom (the probability of the symptoms independent of illness) is actually not very interesting to us, since it is the same for all illnesses.

It makes intuitive sense...

$$P(illness|symptoms) \propto P(symptoms|illness)P(illness)$$

- If the likelihood of symptoms given a certain illness is high, this will increase the posterior probability of that illness
- If the prior probability of a certain illness is high, this will increase the posterior probability of that illness
- If a particular illness has low prior probability, we need some really convincing evidence to make us believe it to be true
    - imagine if your friend had pneumonia, high fever, shortness of breath

Errr... hello... isn't this a course about language?

- In the medical example, we were trying to use evidence provided by symptoms to infer what underlying illness your friend had

- What if you aren't a medic, but a child hearing utterances from a parent and learning a language? You are trying to use evidence provided by utterances to infer what grammar your parent has in their head

  illnesses = languages
  symptoms = utterances
  prior for each illness = **bias in favour of particular languages**

- An ideal language learner will estimate the posterior probability of each possible language given the utterances heard

- Children probably don't calculate sums in their head while learning, but if their learning process is sensible, we can characterise it this way

Bayesian language learning

- Evaluate hypotheses about language given some prior bias (perhaps provided by your biology?) and the data that you've heard

- You want to know the **posterior** but all you have direct access to is the **prior** and the **likelihood** (assuming you know how sentences are produced from a given model of language)

- Bayes' rule provides the solution:

$$P(h|d) = \frac{P(d|h)P(h)}{P(d)}$$

Bayesian language learning

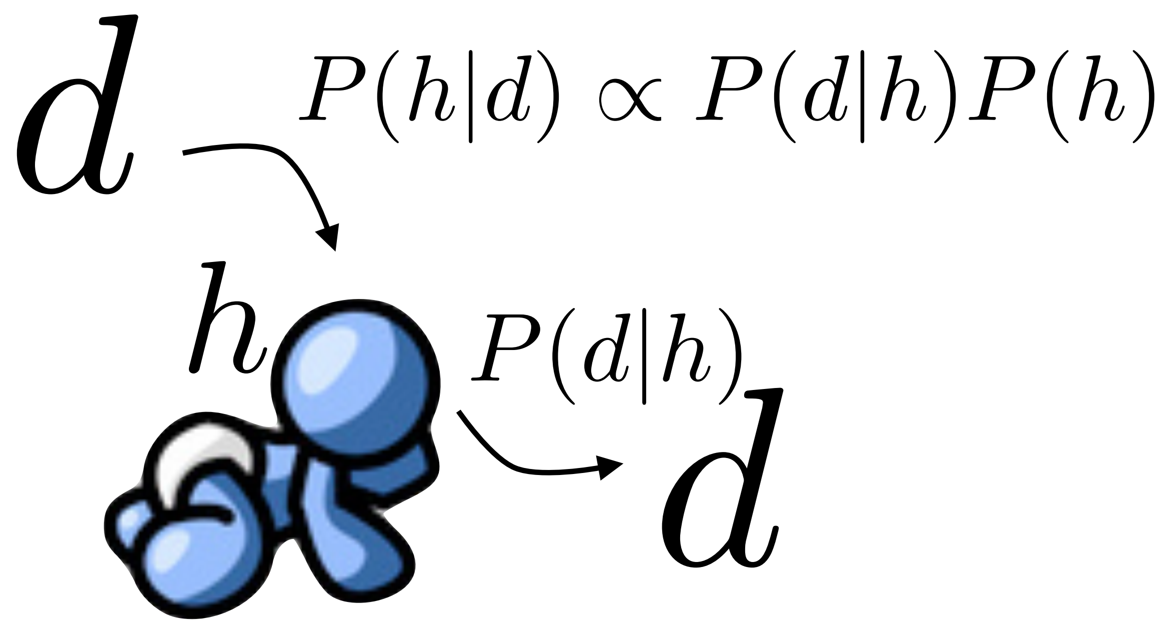
- Evaluate hypotheses about language given some prior bias (perhaps provided by your biology?) and the data that you've heard

- You want to know the **posterior** but all you have direct access to is the **prior** and the **likelihood** (assuming you know how sentences are produced from a given model of language)

- Bayes' rule provides the solution:

$$P(h|d) \propto P(d|h)P(h)$$

Bayesian language learning

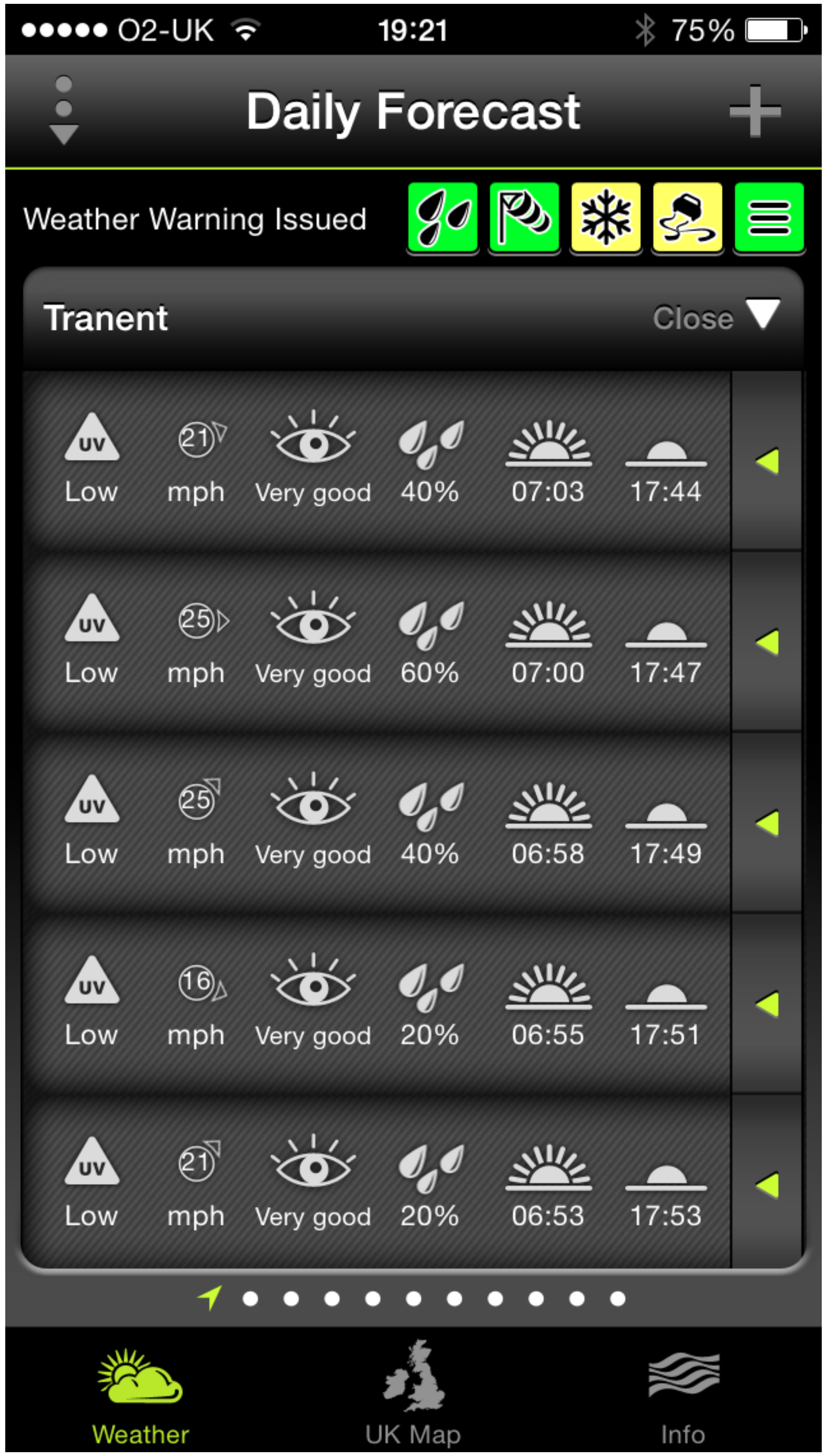


How to put numbers on uncertainty: probability

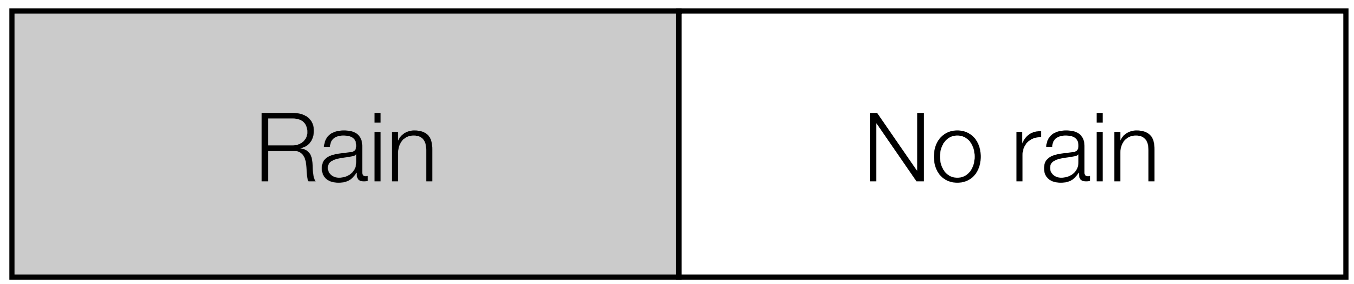
$$P(h|d) \propto P(d|h)P(h)$$

- Bayes Rule is about a relationship between probabilities. What are probabilities?

- How likely is it to rain tomorrow?



Chance of rain



- It will definitely either rain or not rain ("precipitate or not precipitate" - whatever) tomorrow: 100% chance of one of those two outcomes

Chance of rain

- It will definitely either rain or not rain ("precipitate or not precipitate" - whatever) tomorrow: 100% chance of one of those two outcomes

- The met office reckons it's slightly more likely to rain than not: estimate a 60% chance of rain

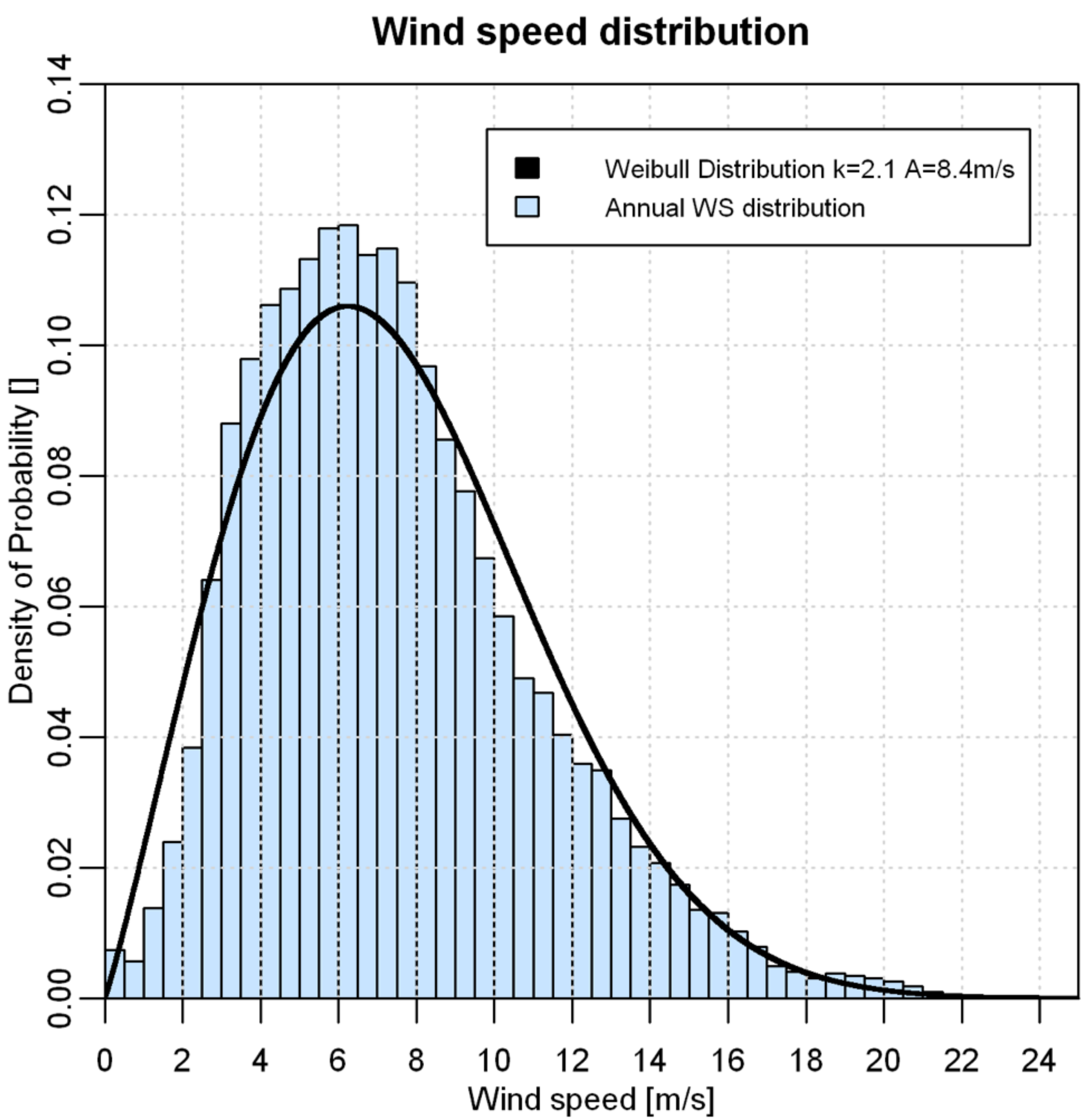
From chance to probability

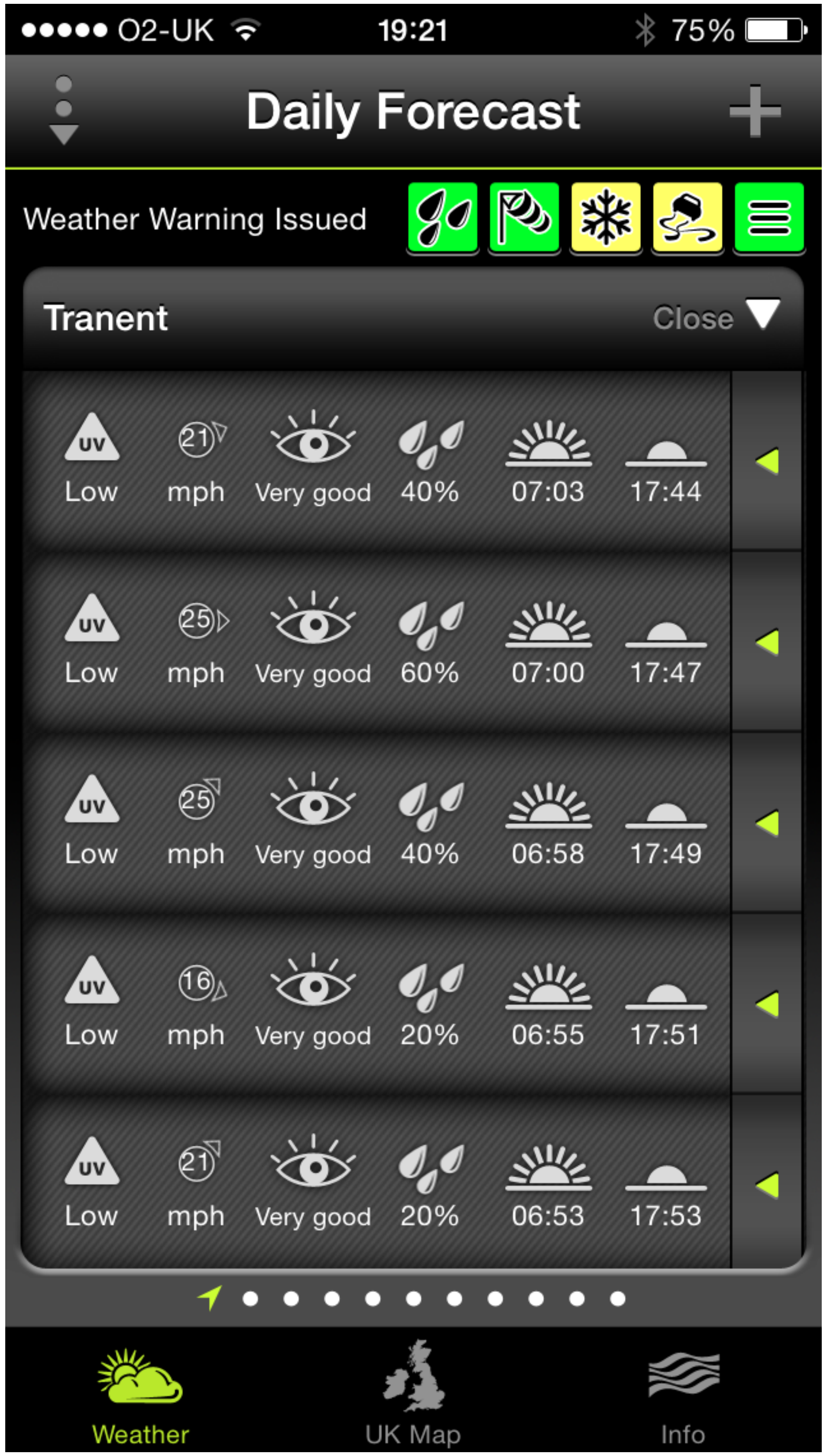
- The Met Office are in fact stating their estimate of the **probability** of it raining tomorrow.
- They express is as a %
- But probabilities are conventionally expressed as a proportion of possible events
  - i.e. they lie between 0 and 1
- “60% chance” corresponds to probability of 60/100 or 6/10 or 0.6

Discrete probability distributions

- A limited set of known possible scenarios
  - Rain, no rain
- We can assign a probability to each event
  - Since the events are mutually exclusive, the probabilities sum to 1

Continuous probability distributions





Summary and next up

$$P(h|d) \propto P(d|h)P(h)$$

- Bayesian learning: a nice simple way to model learning
- Make the bias of learners beautifully explicit
- Involves probabilities:
  - For each possible hypothesis, what is its prior probability? What is the likelihood of the data under that hypothesis?
  - For each possible language, what is its prior probability? What is the likelihood of the linguistic data if people are using that language?
- Tomorrow's lecture: a linguistic case study, and *iterated* Bayesian learning
- Tuesday: lab on iterated Bayesian learning
- Next Thursday: Jennifer Culbertson on word order universals

Reading!

Stone, J. (2013). Bayes’ Rule: A tutorial introduction to Bayesian Analysis. Chapter 1.