Simulating Language
1: why simulate language?

Simon Kirby
simon@ling.ed.ac.uk

Some important questions...

What is the goal of linguistics?

To answer the question:
Why is language the way it is?

OK... but what should we do to approach this question?

The evolutionary approach to language

We answer the **why question** by posing a **how question.**

We can figure out **why** is language the way it is, by understanding **how** it came to be that way.

This course is about work over the past 20 years or so, much of it carried out here in Edinburgh, trying to tackle this **how** question.

The **processes** of language evolution

Answering our how question means figuring out the processes that underpin language evolution.

Processes that shape language:

Language **use**

Language **learning**

INDIVIDUAL

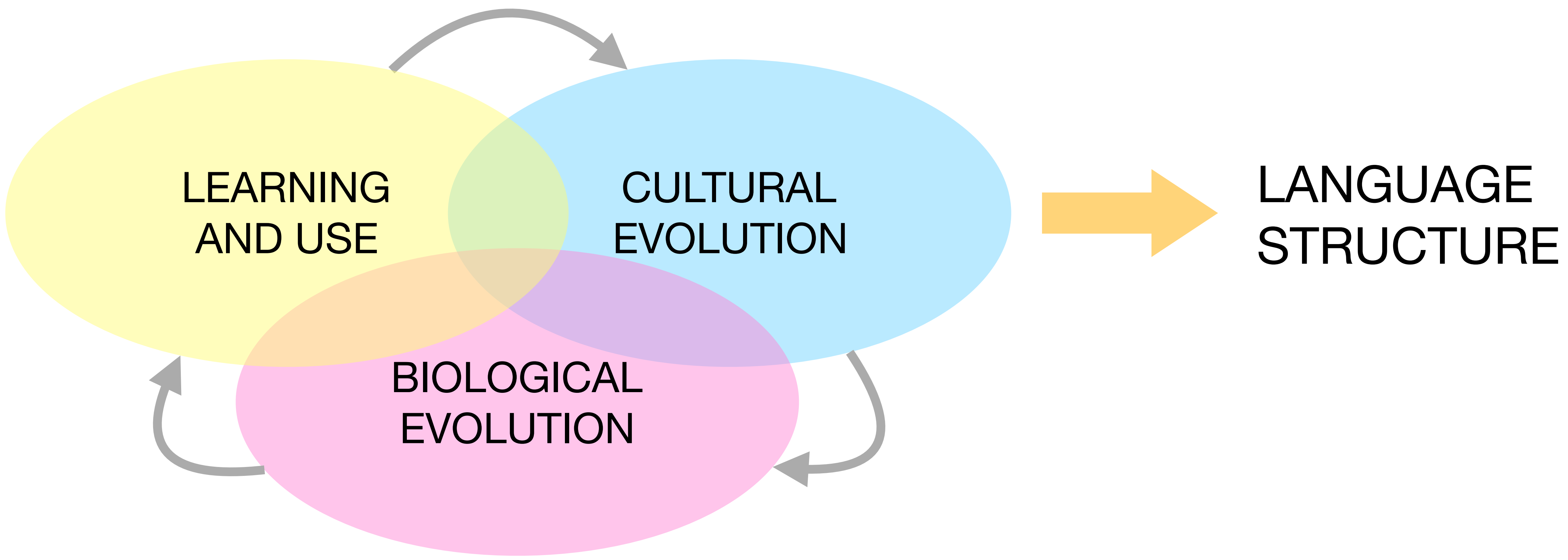
Language **change** (cultural evolution)

Language **evolution** (biological evolution)

POPULATION

Interacting processes

What I'm going to try and convince you of in this course...
You can't really separate any of these processes!



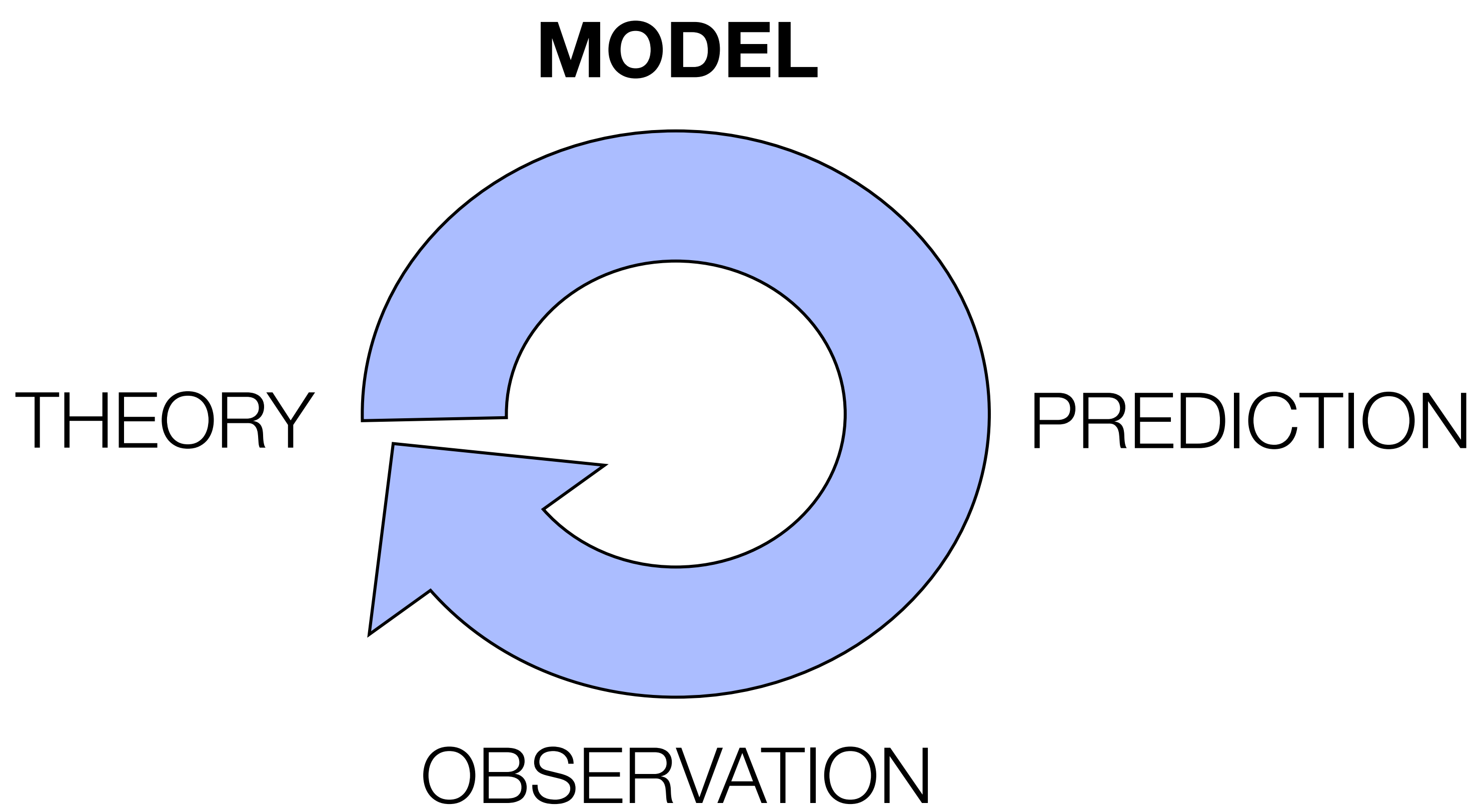
So... how on Earth do we study this?!
We are going to build **models**

What is a model?

- A miniature version of the system you are interested in
- Gains
  - Simplicity
  - Control
  - Ease of observation
- Losses: ?

What is a model for?

- One view:

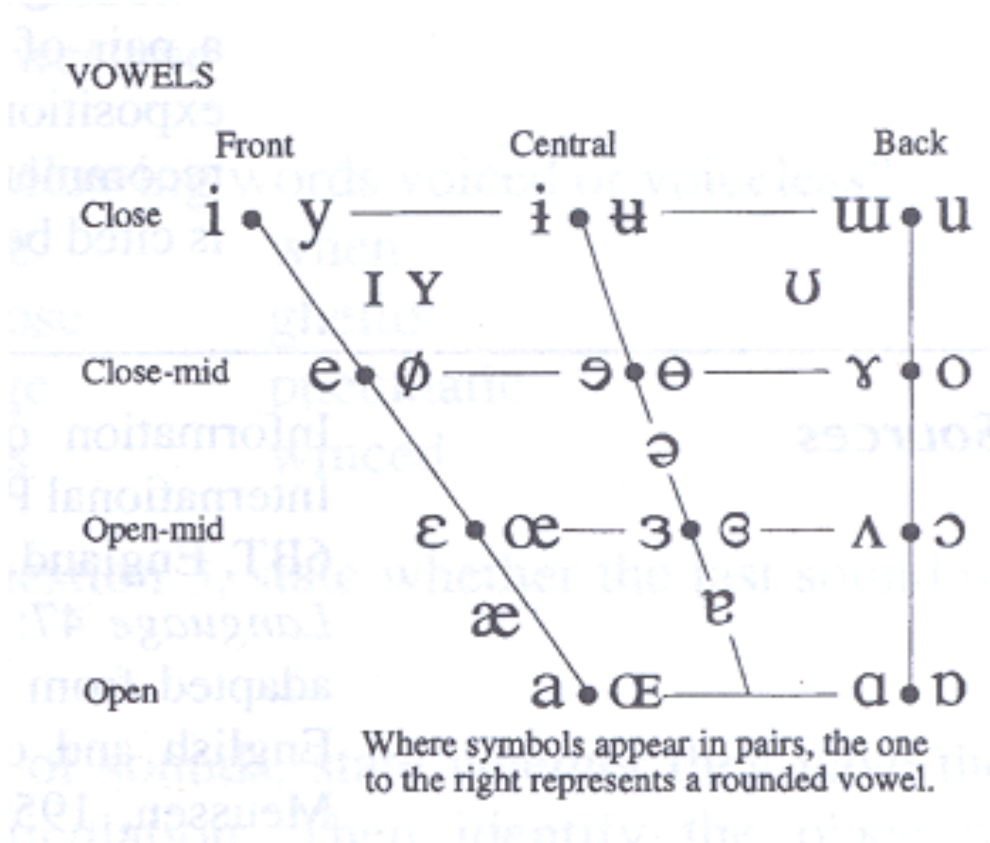


- We use models when we can’t be sure what our theories predict

- Especially useful when dealing with *complex systems*

A simple example

- Vowels exist in a “space”



- Only some patterns arise cross linguistically.
  e.g. Vowel space seems to be symmetrically filled.

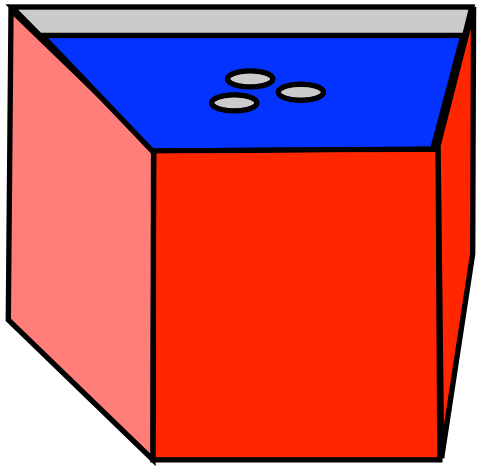
  Why might such universal patterns exist?

The need for a theory

- Possible theory:
  - **Vowels tend to avoid being close to each other in order to maintain perceptual distinctiveness.**
- How do we tell if our theory is correct?

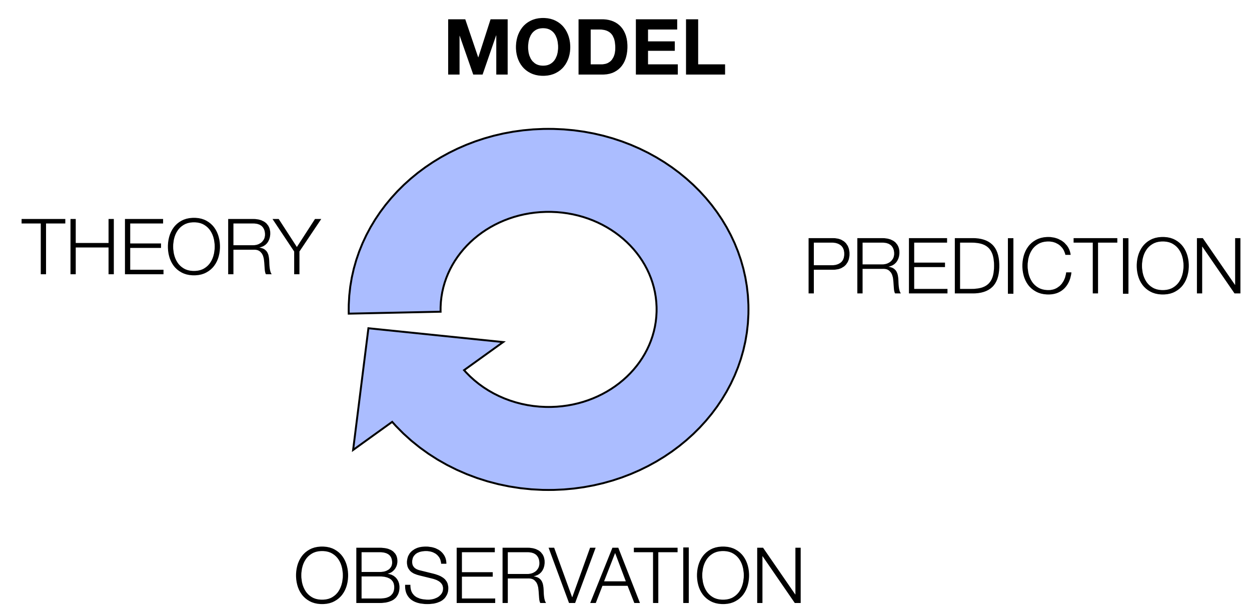
The need for a model

- We need some way of generating predictions from the theory which can be compared with the real data.
- A physical model (Liljencrants and Lindblom 1972): a tank of water, some corks with magnets attached.



- From this model, predicted patterns can be compared with cross-linguistic data

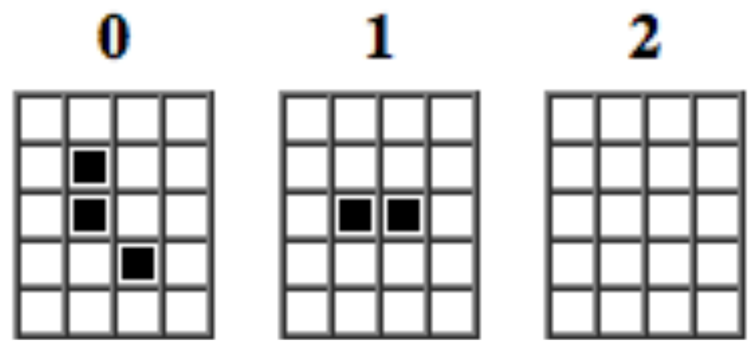
Revisited: What is a model for?

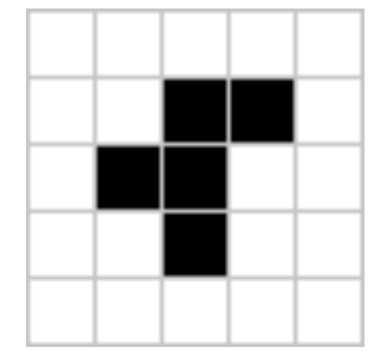


- An alternative / complementary approach: models as tools for understanding

"Predictions are not the pinnacle of science. They are useful, especially for falsifying theories. However, predicting can't be a model's *only* purpose. ... surely the *insights* offered by a model are at least as important as its *predictions*: they help in understanding things by playing with them." (Sigmund, 1995, *Games of Life*, p. 4)

Life





- A ludicrously simple model
- Any cell with 3 live neighbours becomes live.
- Any live cell with 2 or 3 live neighbours stays alive
- Enormously complex behaviours!

The tension between realism and simplicity

- **Question for you: how close to reality should models be?**

  A: The model should leave nothing out, it should be as close as possible to reality

  B: The model should leave lots of stuff out, it should be as simple as possible

My opinion

- What will we learn by building something as complex as the real thing?
- Simpler is better: easier to build, easier to use, better insights
- **Build your model to have as little extra in it that isn't part of your theory**
- We can come back to this

Why use computers for modelling?

- What if:
  - your theory is too difficult to understand simply through verbal argument, or introspection?
  - or a physical model cannot be constructed?
  - or a mathematical model is too difficult (or impossible) to construct?
- Particularly difficult problems involve **dynamic interactions.** For example:
  - a child's knowledge changing as she responds to hearing thousands of words
  - people interacting in groups over thousands of years
  - communicating organisms evolving over millennia

Computational modelling is the solution

- Computers are very good for models of many interacting components
- This has proved particularly valuable in allowing us to build a fundamentally *evolutionary* approach to understanding language
- In this course, we will be building and playing with models to tackle questions like:
    - How do innate signalling systems evolve?
    - How are communication systems shaped by cultural evolution?
    - Where do grammatical generalisations come from?
    - What do we mean when we say language is innate?

Outline of the course

- Half labs, half lectures. Details on website: github.com/smkirby/simlang

- Labs are the primary source of feedback throughout the course. Lots of opportunity for one-to-one help and discussion.

- Rough order of topics:

  Innate vocabularies
  Modelling populations
  Evolution
  Learning
  Cultural transmission
  Bias and innateness
  Syntactic universals
  Co-evolution of genes and culture

This is a **practical** course: lots of time in the lab, working with simulations

- You do not need to know how to program, but you do need not to be scared of computers, and willing to try things out

- We will be working in a simplified subset of **Python**

  We will supply the code for the practicals, but you will need to modify it to carry out the tasks on the interactive notebooks. You can do this from anywhere you have access to a web-browser. (You'll get a log in to a special service called **noteable** that allows this.)

  This isn't a programming course: we aren't going to teach you how to program, but we will teach you just enough to understand and use some simple models we provide. You will have to meet us half way: you'll get on much better if you get your hands dirty and try things out.

Lab classes: 2pm, 3pm or 4pm

- Labs take place in Appleton Tower 3.09
- We need to split up into three groups!
- Doodle poll here: https://tinyurl.com/yc96hug7
- Please mark your preferences on the website, but note that we really need to spread the class over the three sessions as evenly as possible. Some re-jigging might be possible once everyone has indicated an initial preference.
- You need to always go to the same slot irrespective of each day (otherwise it’ll be just too confusing!)

Assessment

- All aspects of the course could be assessed (i.e. every lecture, every lab)
- Two take-home “exams”. Each consists of short answer questions.
- Schedule of assessment release, deadlines, feedback dates available on the course webpage
- For UG students, each worth 50%. For PG students, second worth 70%.

Next up

- Thursday and Friday: your first labs. Python basics.

- Next Monday: how to model *signalling* behaviour in animals. From next week, I want everyone to work through practicals before coming to the labs

- Readings will appear on the Learn page. Sometimes before the lecture, mostly after. There won't be many - focus is on keeping up with the labs.

- Finally, my top tip for this course: **don't miss any lectures or labs!** It'll be very tricky to keep up if you skip any sessions because everything builds on the previous step.