The probabilistic Language of Thought

Or: The pLoT thickens…

- *The Language of Thought: computational cognitive science approaches to category learning*
- Who: Fausto Carcassi
- When: Sommer semester 2022

Refresher: Bayesian update, MCMC

- Repetition solidifies things, so let's do some MCMC by hand!
- Unknown: Temperature
- Prior over temperature: Normal with μ 20.0 and σ 5.0
- Data: Two pretty inaccurate thermometers, saying 34.5 and 33.1
- Thermometer error distribution: Normal with μ 0. and σ 1.
- Let's manually take some posterior samples!

Refresher: Lambda calculus

- Let’s write a lambda expression that
  - Characterizes the set of objects that are red and square (in first order logic, with RED and SQUARE as predicates)
  - Takes an object and returns the set of properties of the object
  - Characterizes the set of objects with exactly two properties

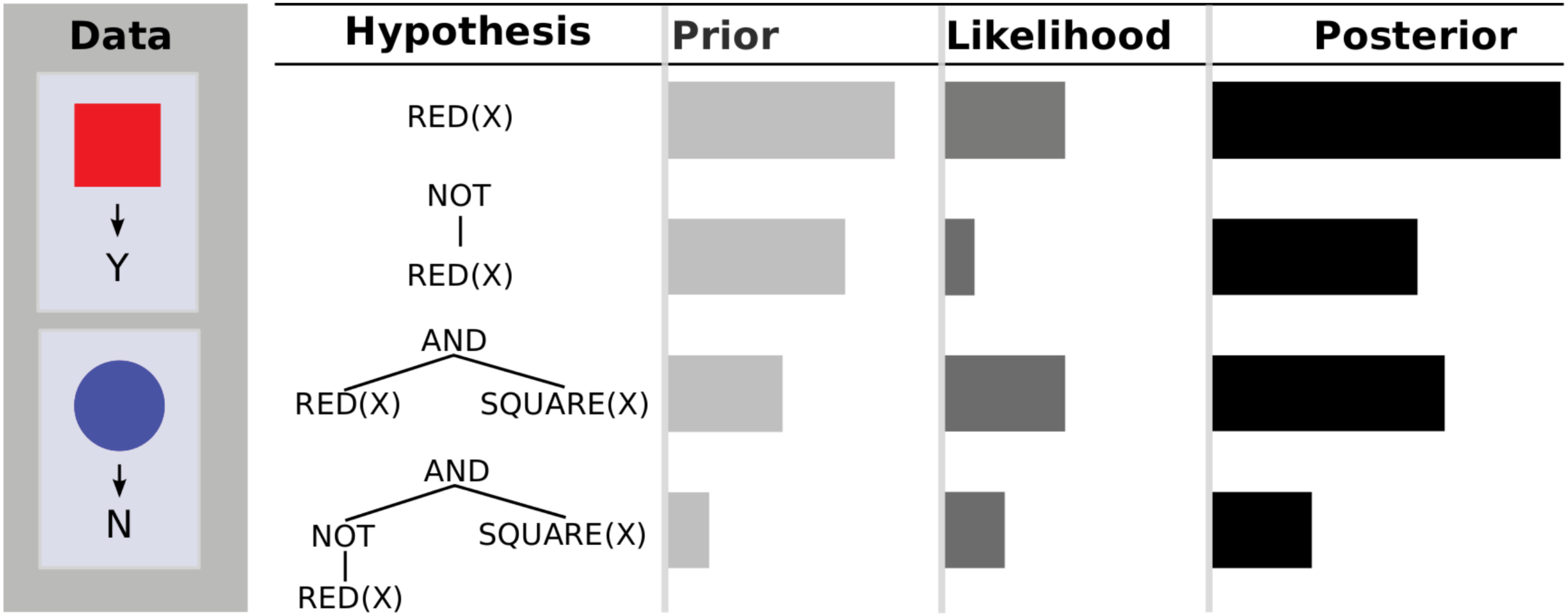
Where we are

- Last week, we have seen how to practically perform Bayesian inference for any given context
  - Namely, approximately and with the MHMC algorithm
  - In practice, this algorithm sometimes doesn't work so well. But it will be fine for our purposes.
  - Piantadosi's other library, *fleet*, implements a better and much faster algorithm (reversible-jump MCMC)
- What we are particularly interested in is doing Bayesian inference on the hypothesis space induced by a PCFG.
- From the exercise you had two weeks ago, you should have a sense of how expressive PCFGs can be. In principle, they can express anything.

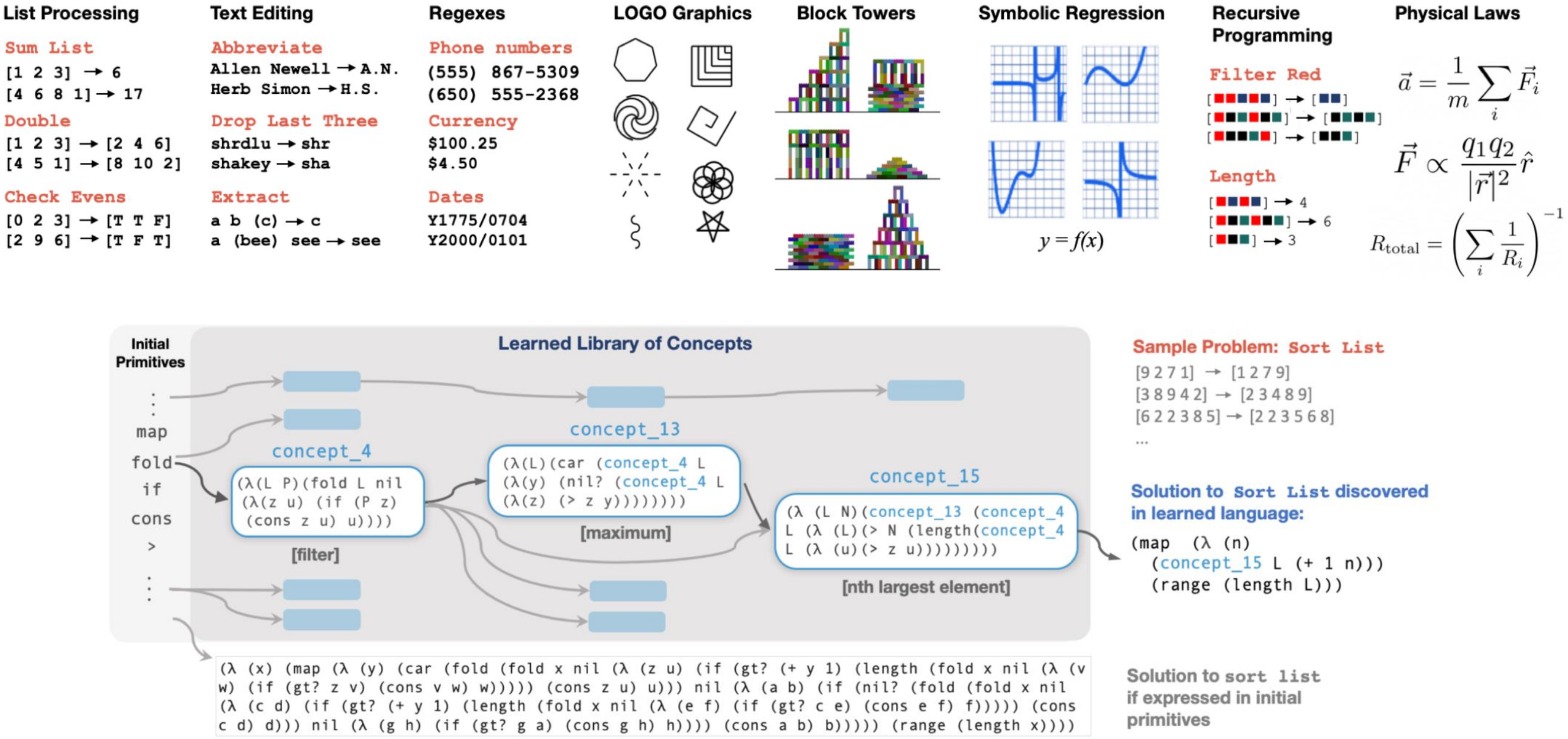
Where we are

- The homework from last week's lab is the last one.
- The other assessment is the final project!
- So please start thinking about what you might want to do for that (some ideas are in the website) and let's discuss it!

The main idea of the pLoT



Application (DreamCoder)



Two approaches to AI

- Historically, there have been two main approaches to AI
- First, a symbolic / logic approach, based on symbol manipulation
  - What are the big advantages and disadvantages of this?
  - Pro: Can model compositionality, deduction, very structured things
  - Contra: Brittle, not robust to noise, can't easily model
- Second, statistical approach
  - Pro: Flexible to noise
  - Contra: Need a lot of hand-specification
- PLoT approach combines these two approaches
  - Representations are modelled symbolically (as opposed to neural nets)
  - But learning happens probabilistically

Nativism and empiricism

- Nativism: Some mental structures / knowledge is innate
  - Pro: Makes little assumptions at the cognitive level
  - Con: Machine learning shows that a lot can be learned from data
- Empiricism: Everything is learned from data
  - Pro: Can make sense of ‘poverty of the stimulus’ arguments
  - Con: Assumes stuff we can’t yet verify
- PLoT finds a balance between the two
  - Most is learned, but the pLoT defines a so-called *structural prior*
  - Specific innate LoT can be studied empirically

Novelty in learning

- Fodor has some prima facie pretty strange views about what’s innate
  - Namely, he thought all concepts are innate (= not-learned)
- Now we can say something sensible about it
  - All concepts are defined in the conceptual system via the PCFG
  - Therefore, concepts aren’t truly learned, in the sense that ‘they’re not there before seeing the data and then suddenly they’re there’
  - Rather, we construct them
- Basically, we assume a small set of primitives that can generate every possible (computable) concept
  - Since λ calculus is Turing-complete, it’s actually enough for everything!

Sensations to concepts

- pLoT inference can infer modality-independent hierarchically-structured representations straight from sensory data
  - “it can specify how local patches combine to form surfaces, how surfaces combine to form parts, how parts combine to form objects, and how objects combine to form scenes”
- Do you see roughly how we could do that?

What can we do with these models?

- Now we have a bunch of techniques. What can we do with them?
- Main: Make some observations, update posterior over sentences in a PCFG
  - This gives us a generative model of some part of the world
- If the PCFG defines functions, it can define higher-order concepts
  - Observations can be input-output pairs ran on the function
  - If the input is e.g. a situation and the output is e.g. a sentence, this can be e.g. a model of language learning
  - We can imagine a system that learns a 'library' of concepts
- Since the PCFG is totally unspecified, and PCFGs can encode anything that's computable, we can learn anything with it.
- Importantly, after we've learned a posterior, we can *inspect* the learned sentence and see what it says in all other cases.

What can we do with these models?

Once we have a generative model, we can do:

- Prediction: predict future world states given the generative model. Predictions can be probabilistic, we can answer counterfactual questions, and we know why they are being made
- Inference: Having a generative model can help us in inferring other aspects of the world. E.g. knowing what color an object is + sensory input can inform us about the lighting conditions.
- Generalization: If we have a generative model, we can transfer the knowledge to new scenarios. E.g. if we learn the weight of a pool ball from seeing some plays, it can help us play future games

Next time: First order logic

- The paper for next time is concerned with inferring categories via inference of expressions in a logical fragment of the pLoT.
- Can you already see how it would go intuitively?
- Do you remember first order logic?

LoT for next time

SimpleBoolean		
START	→	*lambda x . BOOL*
BOOL	→	*(and BOOL BOOL)*
		(or BOOL BOOL)
		(not BOOL)
		true
		false
BOOL	→	*(F OBJECT)*
OBJECT	→	*x*
F	→	COLOR
		SHAPE
		SIZE
COLOR	→	*blue?*
		green?
		yellow?
SHAPE	→	*circle?*
		rectangle?
		triangle?
SIZE	→	*size1?*
		size2?
		size3?

NAND		
START	→	*lambda x . BOOL*
BOOL	→	*(nand BOOL BOOL)*
		true
		false
BOOL	→	*(F OBJECT)*
OBJECT	→	*x*
F	→	COLOR
		SHAPE
		SIZE
COLOR	→	*blue?*
		green?
		yellow?
SHAPE	→	*circle?*
		rectangle?
		triangle?
SIZE	→	*size1?*
		size2?
		size3?

- Aim: define categories
- Based on properties
- Different LoTs give us different inductive biases
- Therefore, based on people's learning patterns, we might be able to infer which LoT they have

Where we’re going

- In the lab this week we will finally learn Piantadosi’s LOTlib3 library
- This will allow us to simply define an arbitrary PCFG, some observations, and run MCMC on it to define a posterior over hypotheses.
- From next week we will then start going through specific case studies to see how the pLoT project has been applied in cognitive science.
- We will alternative reading a paper that implements a pLoT model and trying to implement it in LOTlib3 (or at least a simplified version)
- We have 4 weeks to do this, so 4 models!
- Topics I have picked (but we can change them if you’d rather do something else): Categories, kinship terms, numerals, sequences

Anything unclear this far?