The Logical Primitives of Thought

Or: Logical LoT for categorization

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

Bayesian update

- Let's do a single step of MHMC with the following grammar:
  - S -> 0 | S+1 | S+2 | S-1
- Likelihood function:
  - $\Phi_{Normal}(observation, \mu = I(sentence), \sigma = 2)$
- And the following observation:
  - 3.
- Suppose the starting hypothesis is:
  - 0+1
- Let's use the tree-regeneration algorithm!

Where we are

- Let week, we have seen some arguments supporting the pLoT approach to cognitive science, and we have started to see how to use Piantadosi's LOTlib3 library.
- Today, we'll look at the first application of all the technical machinery we have covered so far, namely the acquisition of logically structured categories
- Paper: Piantadosi et al (2016) The Logical Primitives of Thought: Empirical Foundations for Compositional Cognitive Models

The problem: Concept learning

- Boolean concepts
  - Concepts that use Boolean operators to connect properties
  - E.g. “object is blue and not square”
  - Shefferstrich (negated conjunction) is enough in principle
  - But more operators allow us to express concepts more compactly
  - And different sets of operators can imply different complexity levels for the same category
- Quantificational concepts
  - Concepts that use
  - E.g. “*There is* another object with the same shape”
  - E.g. “*Every* other object with the same shape is blue”

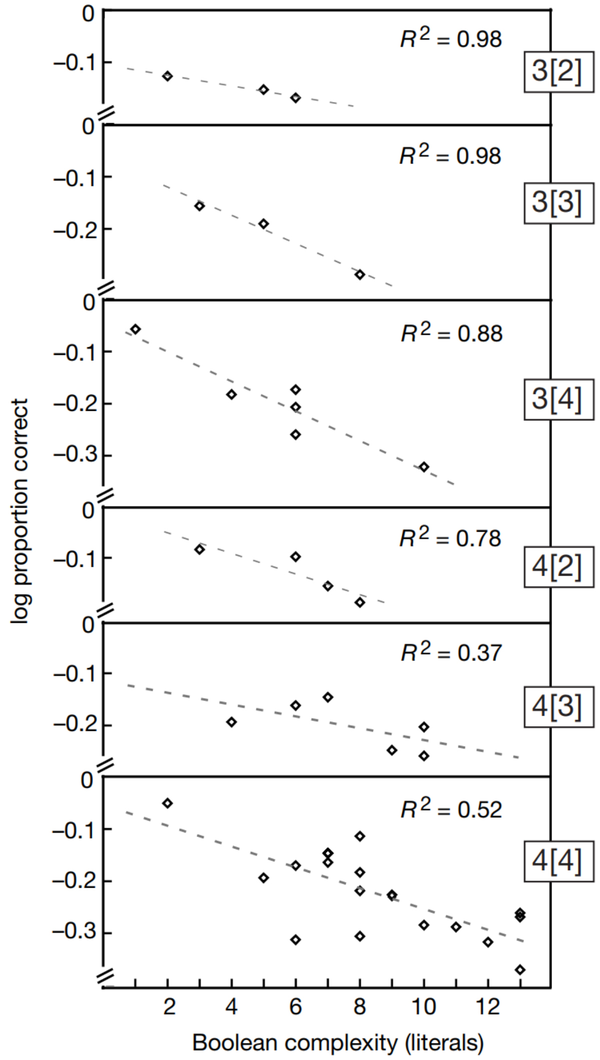
Some background: Feldman (2000)

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

- Feldman (2000) is a very foundational paper for the LoT field, but didn’t age very well.
- Feldman showed participants a bunch of ‘amoebas’ with simple binary features (shape of the nuclei, size of the nuclei, shading of the nuclei and number of nuclei)
- Participants saw a bunch of them and had to learn to identify a ‘new species of amoeba’.
- First a random Boolean concept was generated, and then participants saw all positive and negative examples on the screen for a fixed duration.
- Finally, the participants saw each object and had to say whether it belongs to the new species or not.

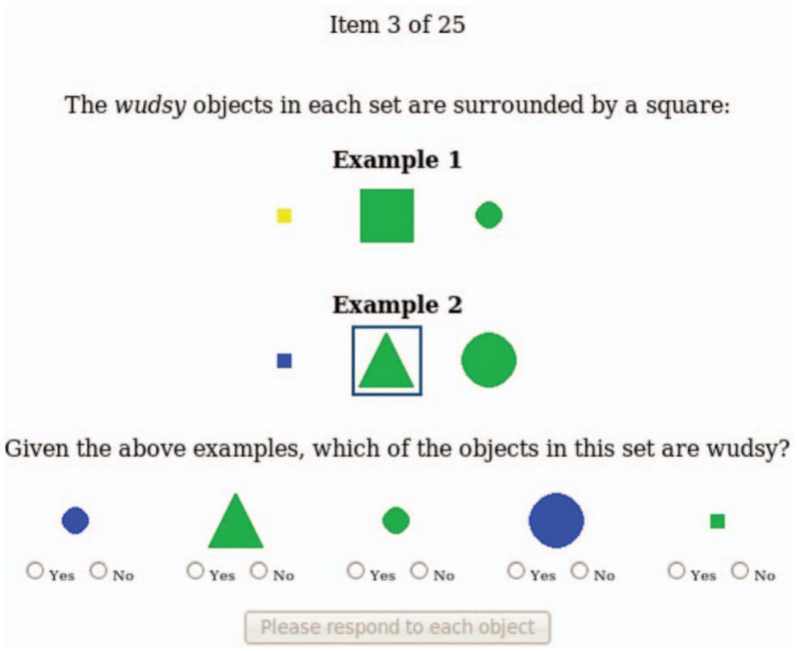
Some background: Feldman (2000)

- What interested Feldman is whether people would struggle categorizing correctly the species that were encoded by more complex concepts in Boolean logic.
- And that's what we see in fact!
- This particular result was disputed in successive literature, but the general approach became very successful.
- Namely: learning manifests something about the complexity of the encoding of different concepts.
- A lot of pLoT literature is an improvement of this!
- In this background, Piantadosi (2016) studies a similar categorization problem.

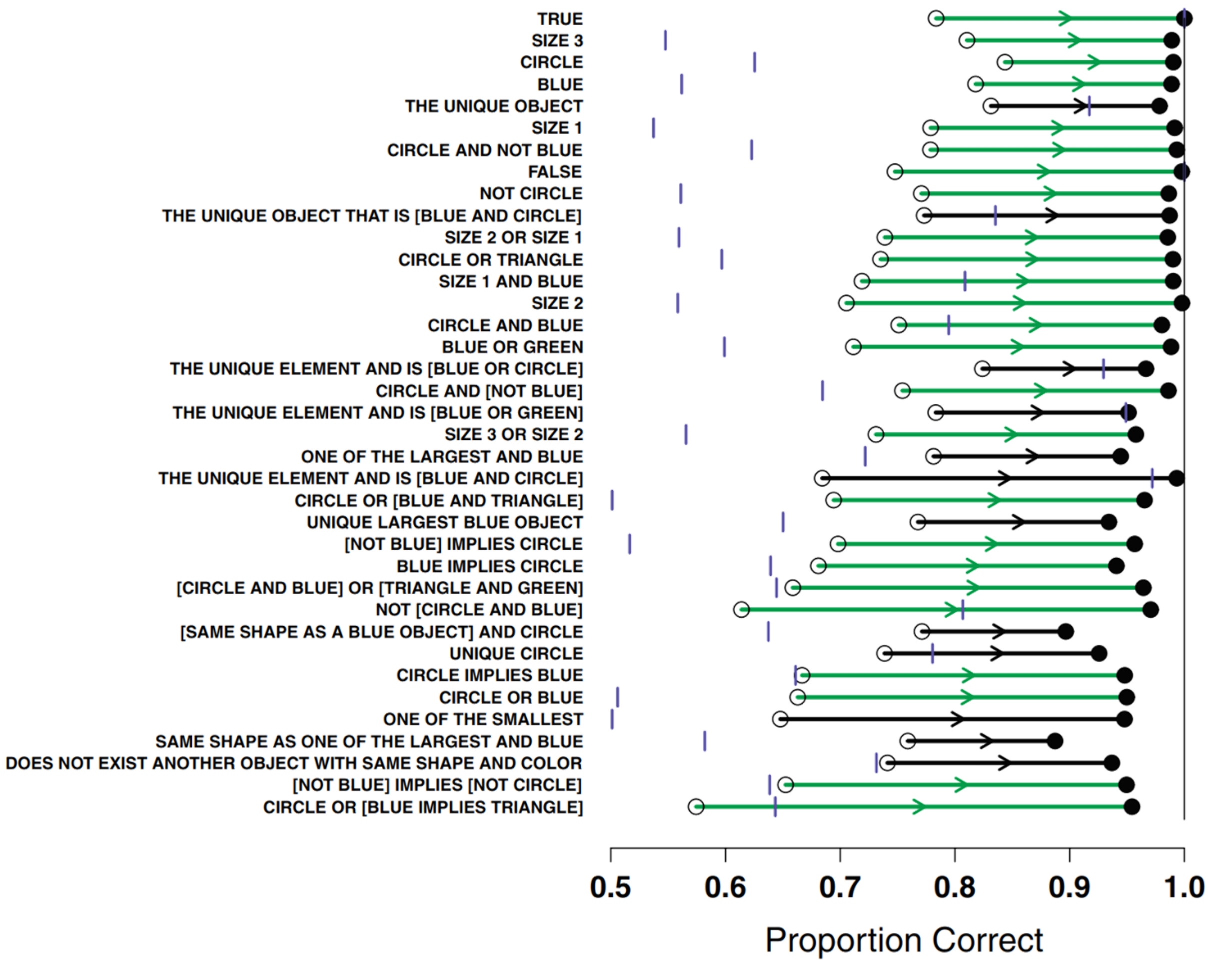


Experiment (Piantadosi et al 2016)

- Participants were told that they had to discover the meaning of wudsy, a word in an alien language.
- They were told that this word applied to some objects in a set, and that whether or not an object was wudsy might depend on what other objects were in the set.
- Participants were shown a set and asked whether each item was wudsy.
- After responding, they were shown the right answers.
- The correctly labeled sets stayed visible on the screen, and participants moved on to the next set.



Experiment



- Objects were:
  - Squares, circles, triangles
  - Green, blue, yellow
  - Three sizes
- 1596 participants, 108 concepts!

Model-free results:

- Top third most easily learned
- White circle: accuracy on first 25%
- Black circle: accuracy on last 25%

Need for a model-based analysis

- Three problems with analyzing this data just looking at accuracy levels:
  - Different concepts have different baseline accuracy which makes it difficult to compare them directly
  - Participants can get high accuracy by learning not the right concept, but a wrong one that agrees with the right one in most cases
  - Data observed by different participants might give different amount of information about the true concept.
- The Bayesian pLoT model allows us to calculate learning curves for specific observed sets and the probability of including each object in the category given learning state!
- Do you see how we could use a Bayesian model to analyse the data?

Bayesian data analysis

- We have looked at how we can do Bayesian analysis when learning about some unknown aspect of the world
- However, note that when we analyse data from cognitive science experiments, what we are interested in is *something about participants we can't see directly*.
- This is perfect for a model of Bayesian inference!
- Suppose we have some model of participants with an unknown free parameter, which tell us the probability that the participant will behave in a certain way: P(behaviour | values of hidden parameters, cognitive model)
- Then, we can use Bayes theorem to find a distribution over the hidden parameter given the experimental data! This is *Bayesian data analysis*.
- This is exactly what they do in the Piantadosi model.
- So let's see what model they develop for participants' behaviour!

Boolean LoTs

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

- Instead of using just one LoT, Piantadosi et al consider a set of possible LoTs that participants might be using the infer the data, e.g.:

SIMPLE-FOL

(SIMPLEBOOLEAN *rules not shown*)

SET	→	*S*
	→	*(non-Xes S)*
BOOL	→	*(forall F SET)*
		(exists F SET)

FOL

(SIMPLEBOOLEAN *rules not shown*)

F	→	*(lambda* x_i *. BOOL)*
SET	→	*S*
	→	*(non-Xes S)*
BOOL	→	*(forall F SET)*
		(exists F SET)
		(size>= OBJECT OBJECT)
		(size> OBJECT OBJECT)
		(equal-size? OBJECT OBJECT)
		(equal-color? OBJECT OBJECT)
		(equal-shape? OBJECT OBJECT)

SIMPLEBOOLEAN

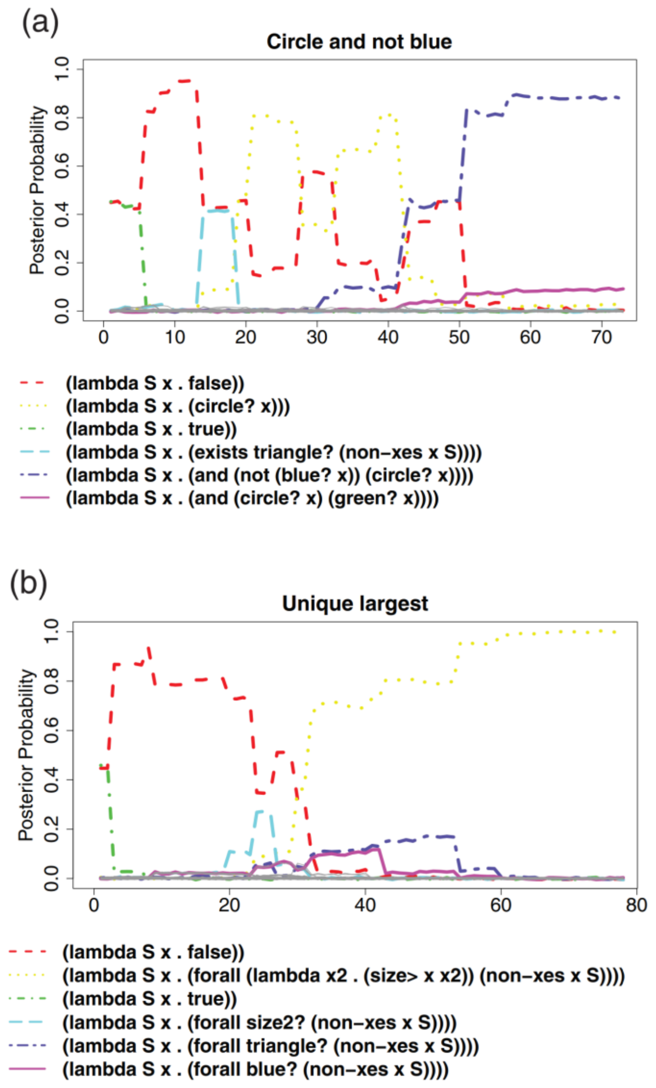
START	→	*lambda x . BOOL*
BOOL	→	*(and BOOL BOOL)*
		(or BOOL BOOL)
		(not BOOL)
		true
		false
BOOL	→	*(F OBJECT)*
OBJECT	→	*x*
F	→	COLOR
		SHAPE
		SIZE
COLOR	→	*blue?*
		green?
		yellow?
SHAPE	→	*circle?*
		rectangle?
		triangle?
SIZE	→	*size1?*
		size2?
		size3?

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?

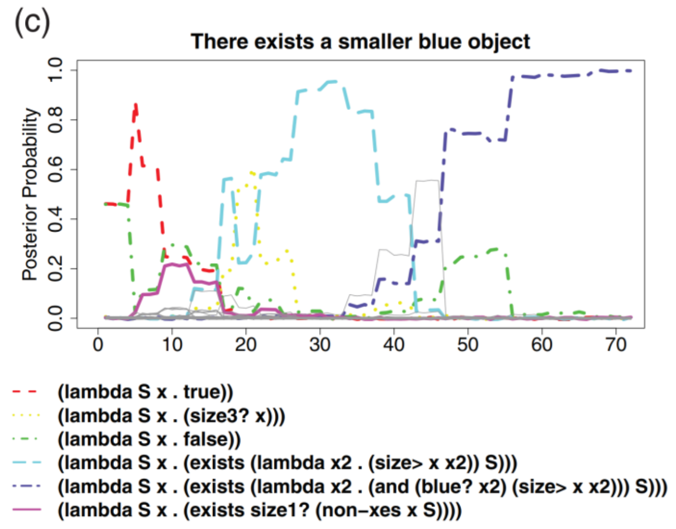
Bayesian model

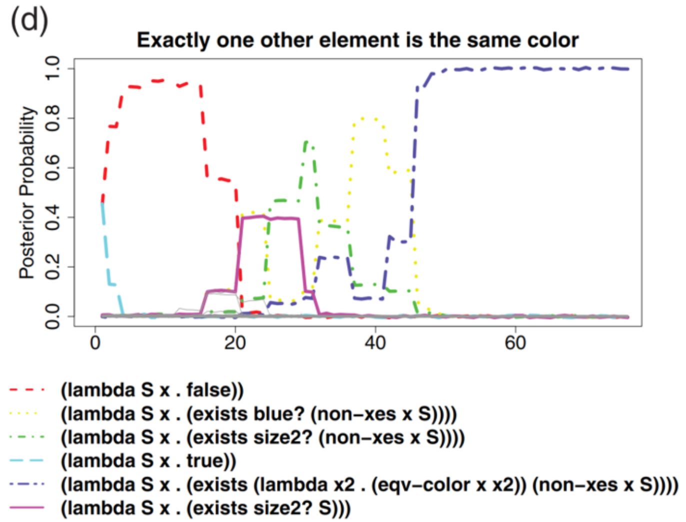
- Prior $P(h \mid G, D_{**})$
  - Hypothesis h
  - Grammar G
  - (Roughly) production probabilities D_{**}
- Likelihood $P(l_i \mid h, s_i, \alpha, \gamma, \beta)$:
  - Probability that set s_i was labelled l_i if h is the true concept
  - Amount of noise α
  - Baseline preference for true responses γ
  - Memory decay β
- E.g., in pics: $\alpha = 0.75,\ \gamma = 0.5,\ \beta = -0.1$



Bayesian model

- Prior $P(h\,|G, D_{**})$
  - Hypothesis h
  - Grammar G
  - (Roughly) production probabilities D_{**}
- Likelihood $P(l_i \mid h, s_i, \alpha, \gamma, \beta)$:
  - Probability that set s_i was labelled l_i if h is the true concept
  - Amount of noise α
  - Baseline preference for true responses γ
  - Memory decay β
- E.g., in pics: $\alpha = 0.75,\ \gamma = 0.5,\ \beta = -0.1$





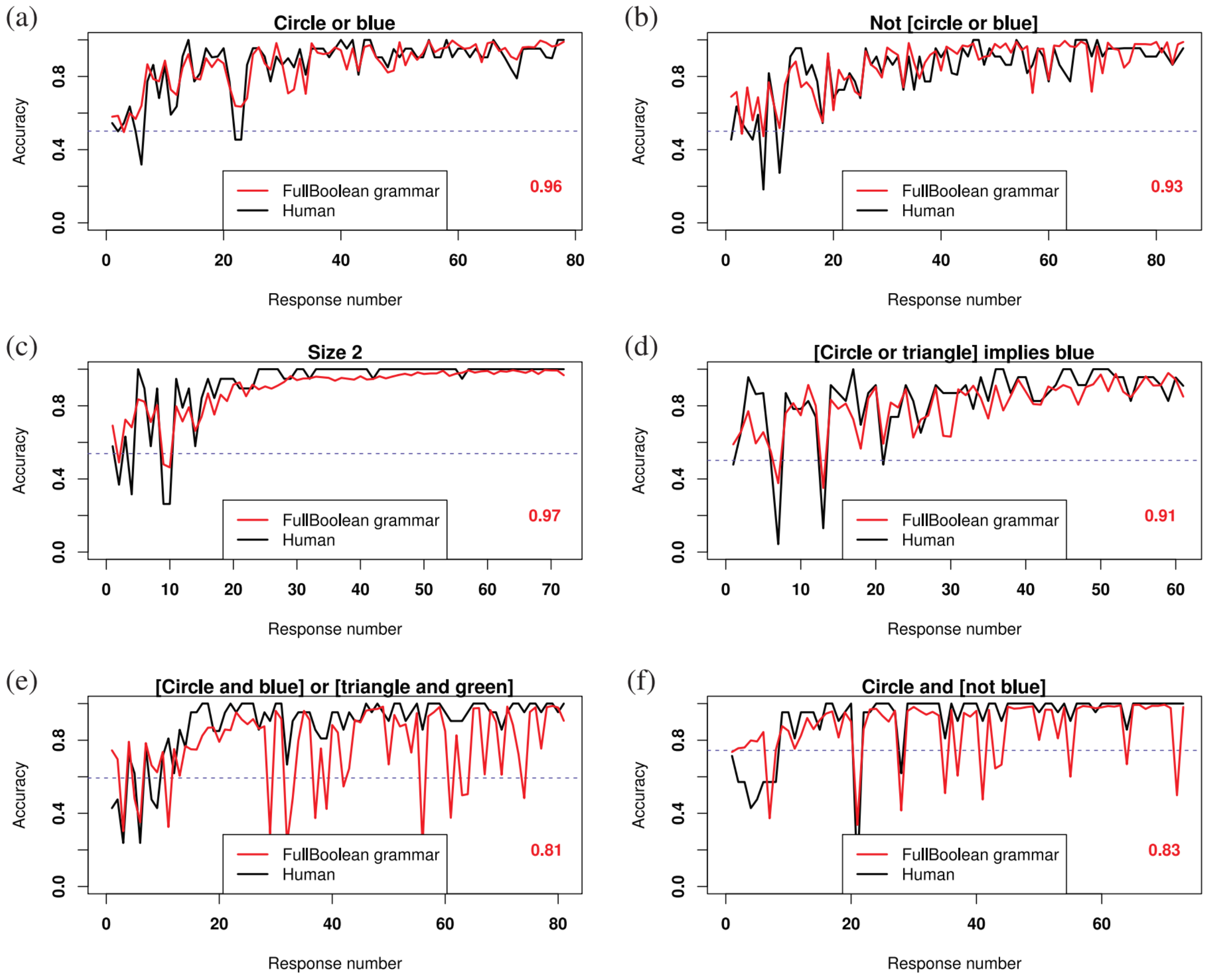
Bayesian model

- Now we have a model of *participants' behaviour*
- So we can do Bayesian inference on the experimental data and calculate a posterior for the unobserved parameters that control participants' categorization behaviour.
- The posterior depends on the grammar that we assume participants are using.
- Piantadosi fits the data with various grammars, e.g. for the purely Boolean:

Language	Description
SIMPLEBOOLEAN	*and*, *or*, *not*, used in any composition.
IMPLICATION	Same as SIMPLEBOOLEAN, but with logical implication (⇒).
BICONDITIONAL	Same as SIMPLEBOOLEAN, but a biconditional operation (⇔).
FULLBOOLEAN	Same as SIMPLEBOOLEAN, but with logical implication (⇒) and biconditional (⇔).
HORNCLAUSE	Expressions must be conjunctions of Horn clauses (e.g., *(implies (and (and a b) c) d)*).
DNF	Expressions are in disjunctive normal form (disjunctions of conjunctions).
CNF	Expressions are in conjunctive normal form (conjunctions of disjunctions).
NAND	The only primitive is *NAND* (not-and).
NOR	The only primitive is *NOR* (not-or).
ONLYFEATURES	No logical connectives; the only hypotheses are primitive features (*red?*, *circle?*, etc).
RESPONSEBIASED	Learners only infer a response bias on *true*/*false*.

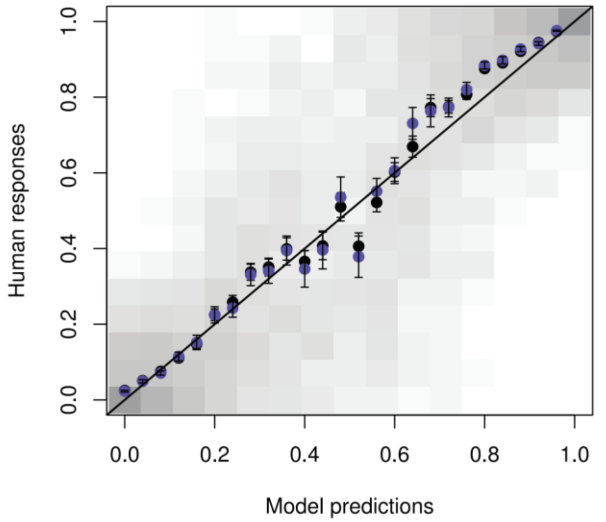
Results

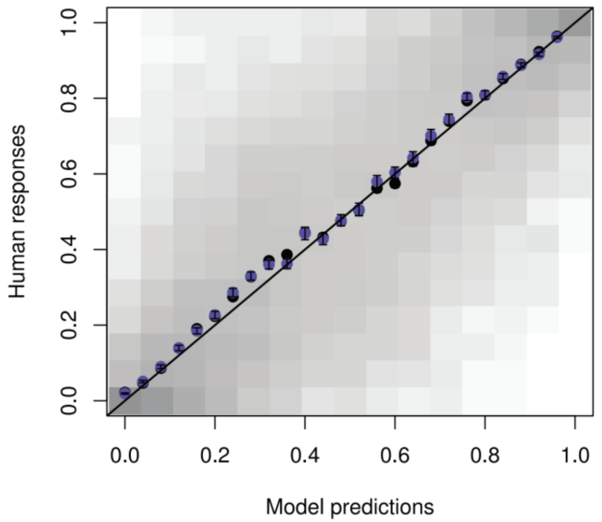
- One crucial question is how well the Bayesian LoT model captures learning behaviour.
- Plot shows FullBoolean for various categories that it can learn vs participant accuracy (red number is R^2.
- (a)-(d) are chosen as good examples, (e)-(f) as bad.



Results

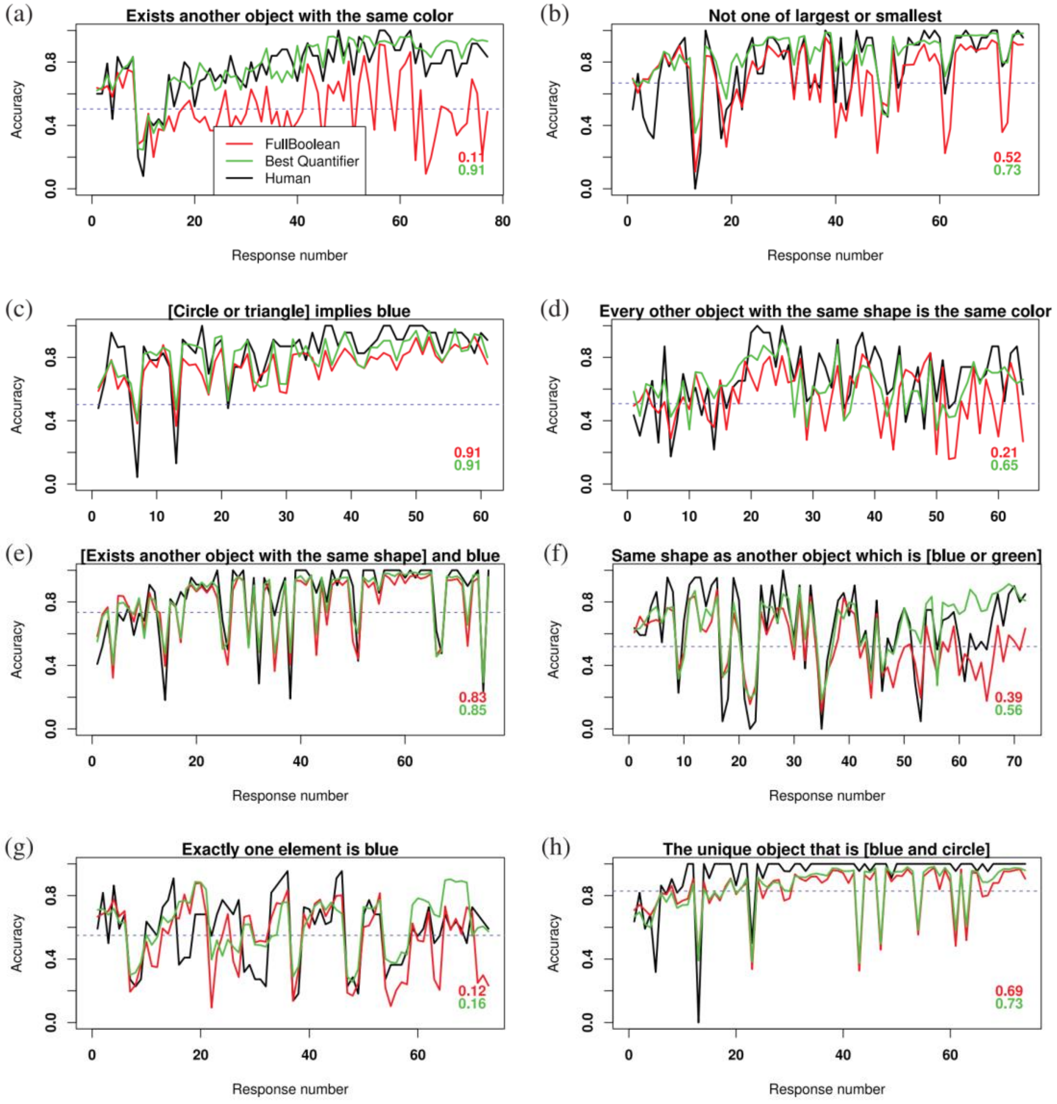
- Really good correlation between model’s predictions and participant’s categorization probabilities in the Boolean case (top fig)
- Even better with quantificational LoTs! (bottom fig)





Results

- A second question is *which* grammar best captures the data.
- It seems clear that LoTs with quantification capture learning patterns much better.
- This indicates the participants are using quantificational means of representing categories!



Where are we now?

- This week we have seen that a LoT model can accurately model human learning of feature-based categories.
- Categorization is a fundamental domain for cognitive science, so this model is quite an impressive achievement.
- In the lab this week we'll finish looking at the introduction to LOTlib3, and if there is time look at an implementation of category learning in LOTlib3.
- Next week we have a choice!
- We can either look at:
  - A paper from last year that concerns the acquisition of *kinship systems* which also looks at kinship systems in different languages.
  - A paper from 2015 on inferring hand-written digits with an LoT