Introduction to modelling of behavioral data

BAMB! 2023 Summer School
Lecture 1A – Alex Hyafil

Life cycle of a modelling project

Our goal for today: cover all the steps involved in a modelling project using a toy model example

✓ Defining a model

✓ Simulating a model

✓ Fitting a model to behavioral data

✓ Validating the model against the data

✓ Comparing various models on a dataset

✓ Checking the validity of the model selection process

Part I
Defining your model
(a 5-step guide)

BAMB! Summer School
Lecture I

1. Start from a question

(about the brain)

One model

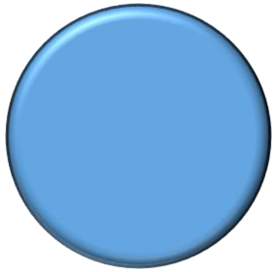
=

One hypothesis

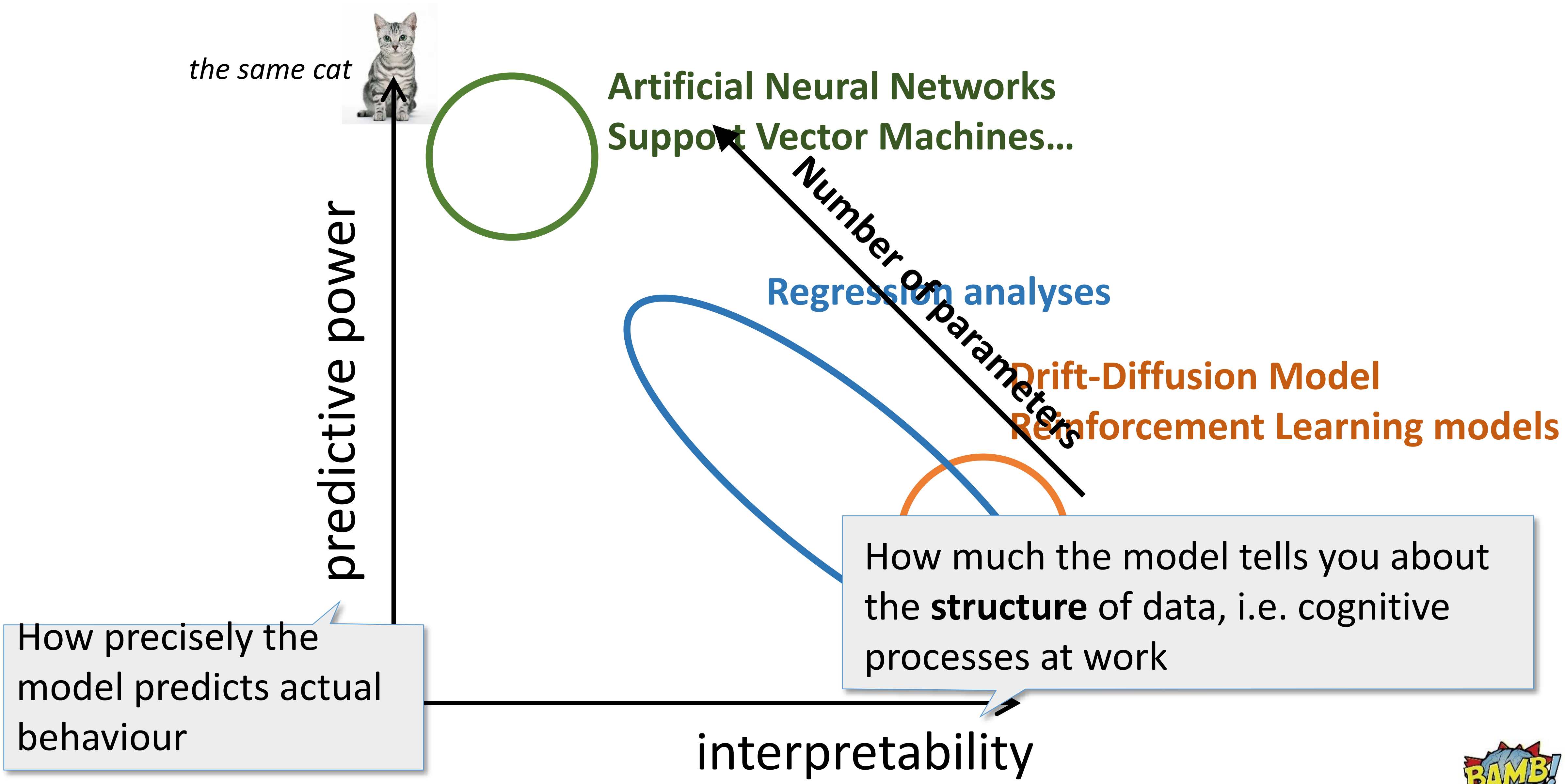
2. Define your modelling objectives

"All models are wrong, some are useful."

George Box, 1976

Is  a good model of  ?

Predictive power vs. interpretability



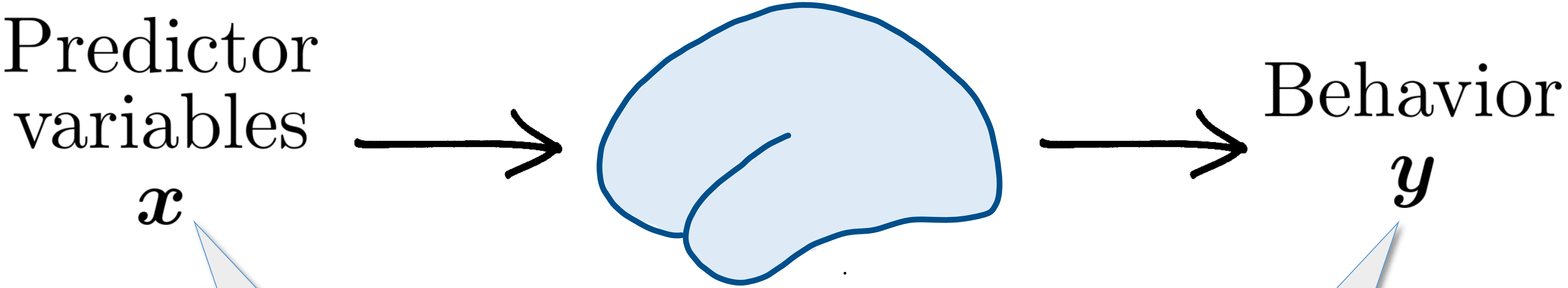
Setting your modelling on firm experimental ground

Modelling does **not** spare looking for an experimental paradigm where models generate different *conceptual* predictions. Otherwise it is unlikely modelling can be of any help.

Modelling as **support** to standard analyses rather instead of standard analyses.

A good modelling won’t save a bad paradigm!

3. Define your input and output variables



Called *independent variable, regressors, covariates, explanatory variables* …:

- Stimuli
- Experimental conditions
- History of stimuli, choices, outcomes, …
- Peripheral marker (SCR, ECG, pupil dilation, …)
- Neural marker (alpha power, ROI activity,…)
- …

able, target variable …:

s

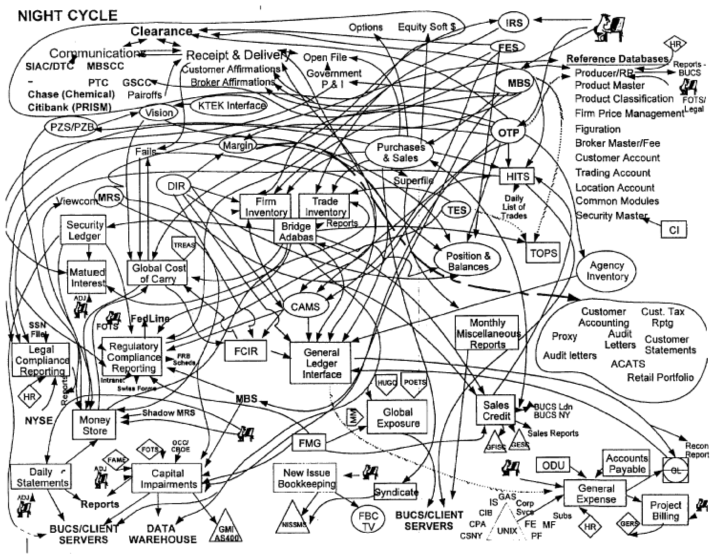
Selecting output variables

- The output variables should be sensitive to the process under study (sure)
- Include more than one variable if they offer complimentary views on the process under study
(e.g choice & RT in speeded perceptual tasks)

input variables = variables of interest + control variables

Should I include this control variable?

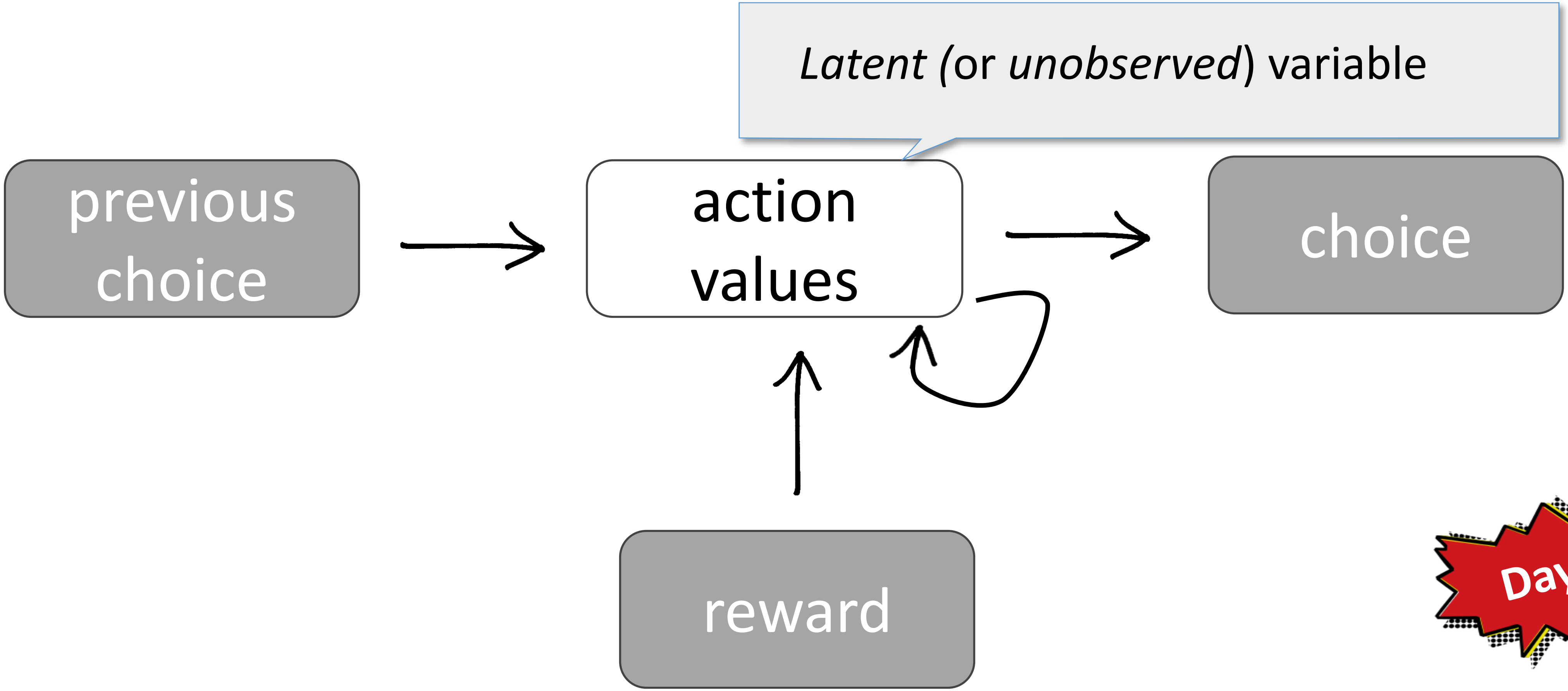
- <u>always</u> when the *control* variable is correlated with the variable of interest
  -> regressing *out* the effect of control regressors

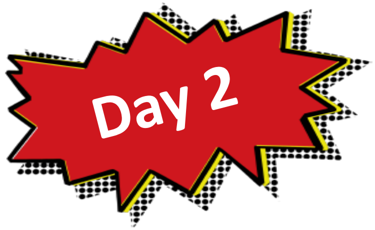


- If it is not correlated with the variable of interest, adding the control variable can have two opposite effects:
  - if it has a strong influence on behavior, more variance is explained so the parameter for the variable of interest will be more precisely evaluated
  - If the influence is weak and the number of trials is low, more regressors may lead to *overfitting*, so poorer estimation

4. Define your model as a box diagram

The diagram **isolates cognitive processes** at play in producing behavior





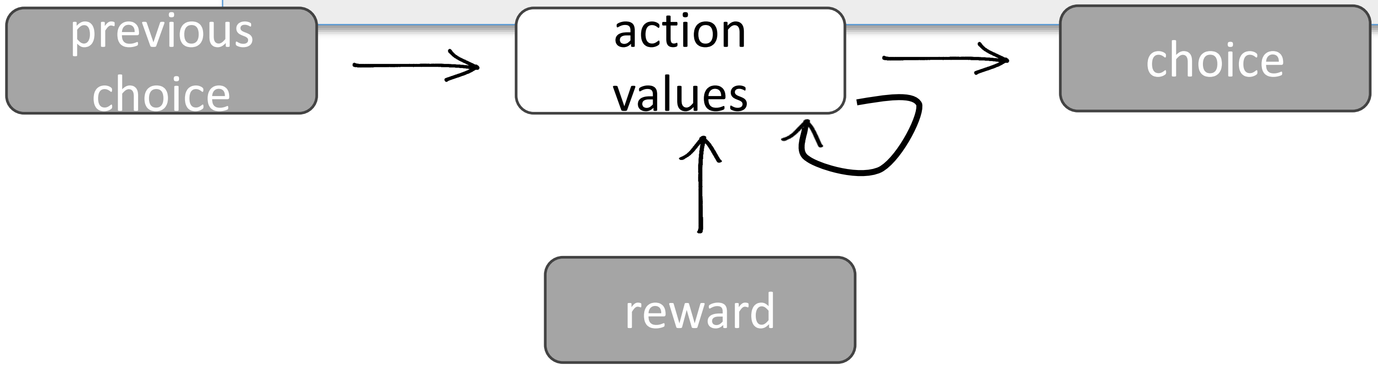
5. Convert the diagram into a statistical model

- We need to formalize our model by moving from a conceptual description to a computational level
- Formalization allows:
  - ***Model estimation***: compare *parameters* in different conditions, populations; infer *latent variable* to look for their neural substrate (e.g. model-based fMRI)
  - ***Model validation and comparison***: can one model account for subject behavior? Better than others?
  - ***Model predictions***: formalization allows to make specific predictions (both qualitative and quantitative) that often are difficult to formulate directly from a conceptual description.

A statistical description of behavior

$$p(\boldsymbol{y}|\boldsymbol{x}) = f_{\mathcal{M}}(\boldsymbol{y}; \boldsymbol{x}, \boldsymbol{\theta})$$

Parameters (properties of neural system), e.g. learning rate, sensitivity to motion, etc.



Turning arrows into (deterministic/stochastic) equations

action values: $V_k^{(t+1)} = V_k^{(t)} + \lambda(reward - V_k^{(t)})$

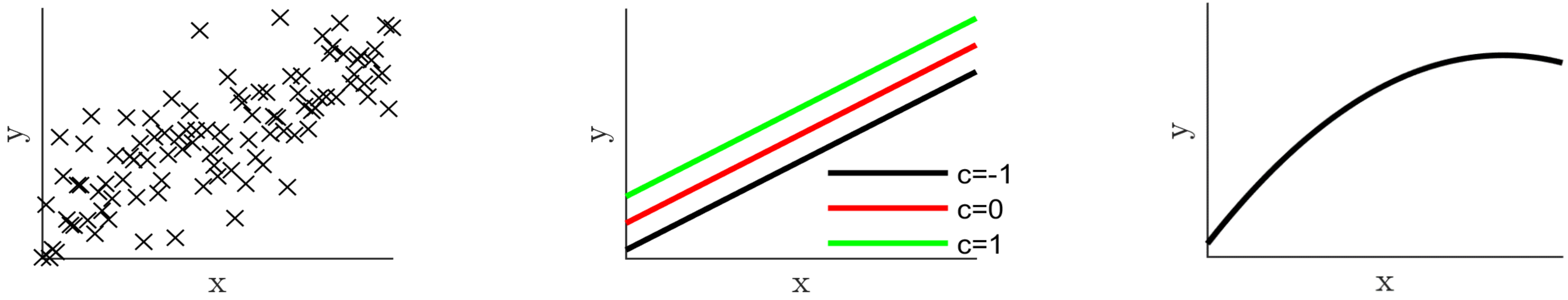
choice: $p(choice = k) = e^{V_k^{(t)}} / \sum_l e^{V_l^{(t)}}$

What’s that noise (in the model)?

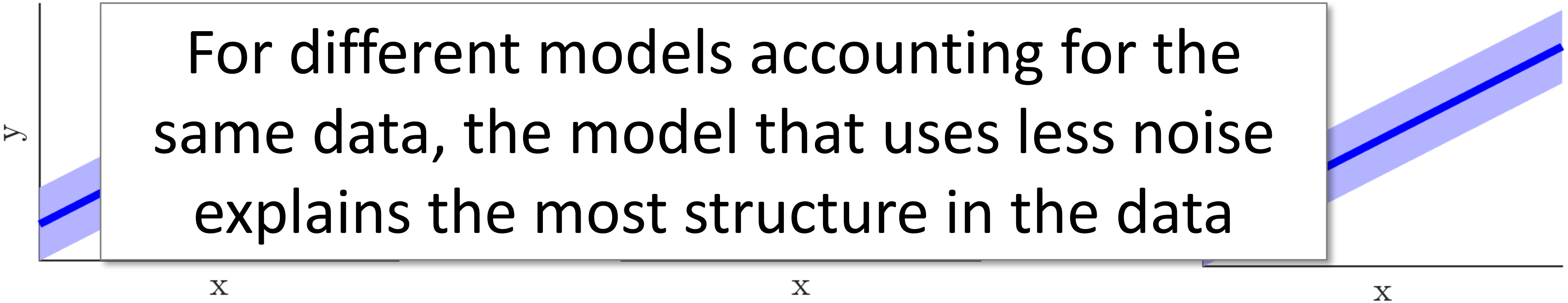
‘Noise’ = all that cannot be accounted for by the deterministic part of the model:

- **Neural noise**: low level variability in neural circuits / behavior (sensory, decision, motor…)
- **Factors not taken into account** in the model (fluctuation in response bias, attention …)
- **Model imperfections**: any systematic bias between the model predicted response and subject behavior that the model cannot accomodate

DATA



MODEL



For different models accounting for the same data, the model that uses less noise explains the most structure in the data

Preparing for tutorial part 1

- Standard 2AFC motion discrimination task (left vs right)
- Stimulus evidence is varied across 9 levels
- In half of trials, brain area X is experimentally deactivated
- A reduction in performance is observed in those trials

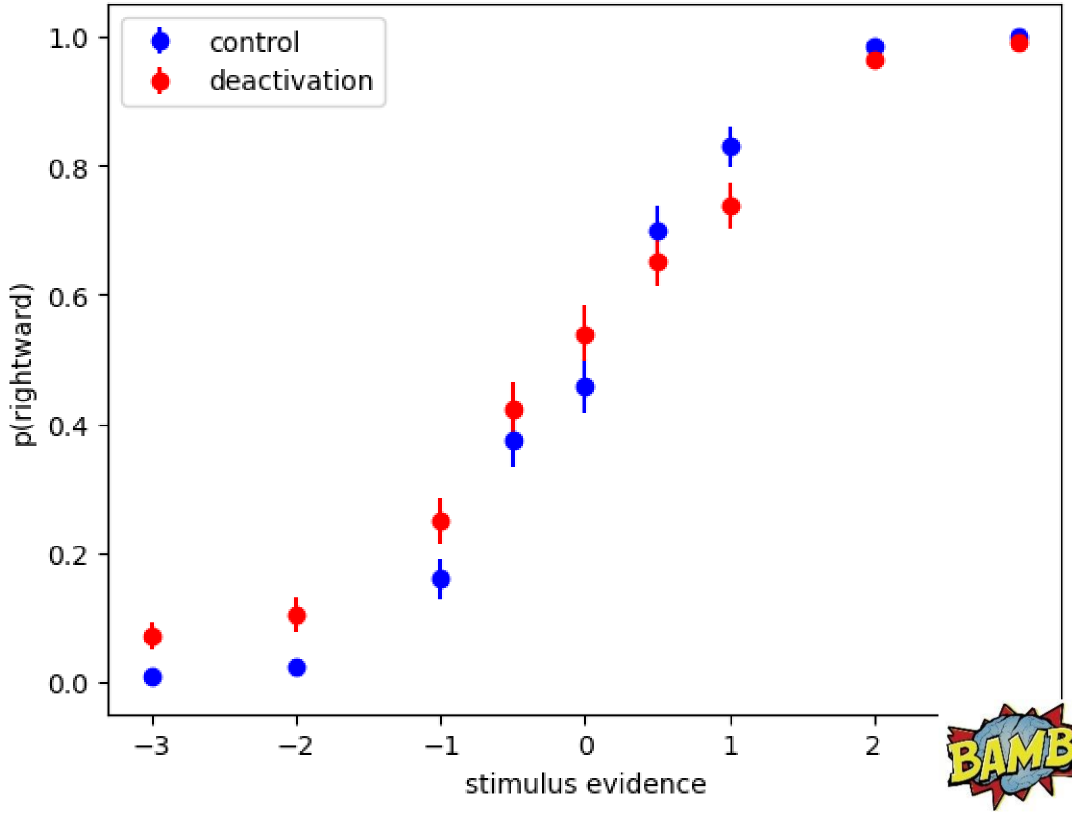
Is it due to deteriorated sensory processing or to sensory-unrelated effects (lapses)?

Defining the models

1. <u>Scientific question:</u>
   "Does deactivating brain area X deteriorate sensory processing or provoke sensory-unrelated effects (lapses)?" -> Two models

2. <u>Modelling objectives</u>:

- which of two models representing each hypothesis better accounts for the data? (*relative criterion*)
- does the favored model reproduces the change in the psychometric curve during deactivation? (*absolute criterion*)

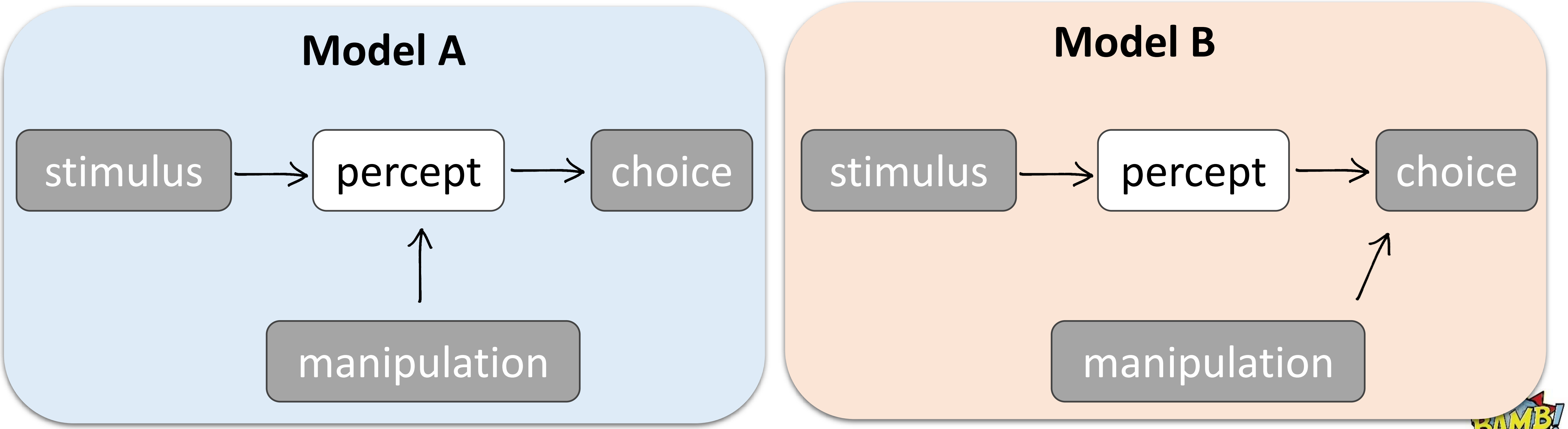


Defining the models (cont'd)

3. <u>Defining input and output variables:</u>

   - output variable: choice (binary)
   - input variables: stimulus evidence, manipulation (*+control variables?*)

4. <u>Defining the models as box diagrams</u>:



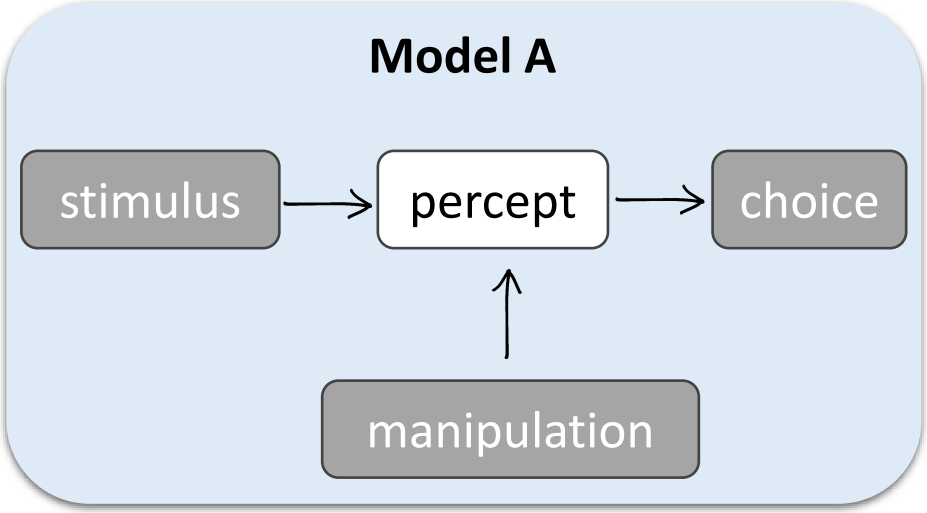
Defining the models (cont'd)

5. <u>Convert model A to statistical model:</u>

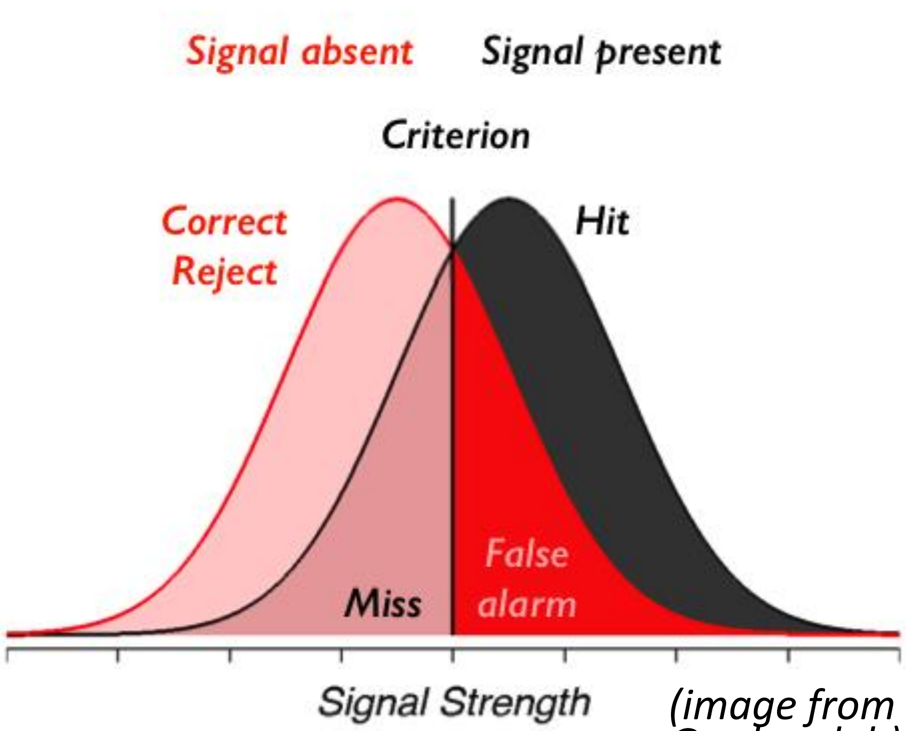
percept: $\hat{s} = s + \epsilon$

ϵ sensory noise drawn from zero-mean Gaussian, variance σ_d^2 in deactivation trials, σ_m^2 in control trials

choice: Right if $\hat{s} > 0$, Left otherwise



Signal detection theory



(image from Gardner lab)

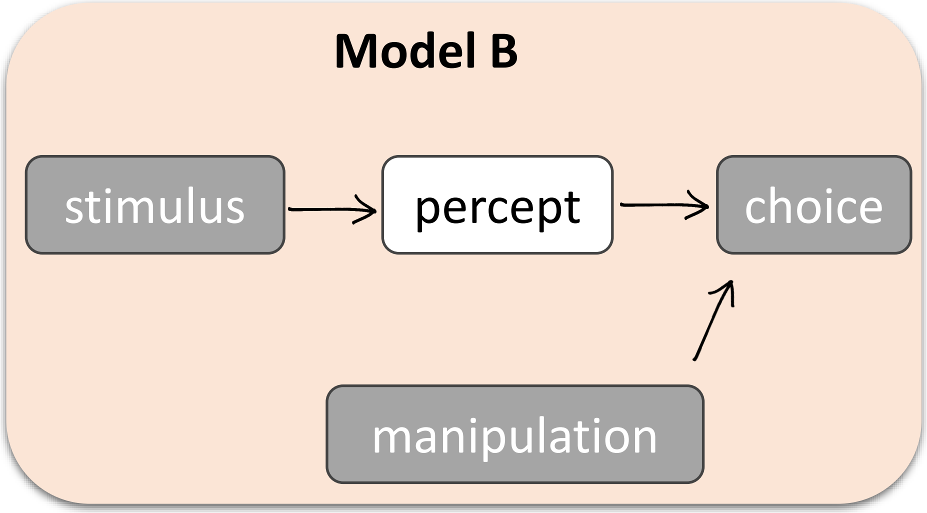
Defining the models (cont’d)

5. <u>Convert model B to statistical model:</u>

percept: $\hat{s} = s + \epsilon$

ϵ sensory noise drawn from zero-mean Gaussian, variance σ_d^2 in all trials

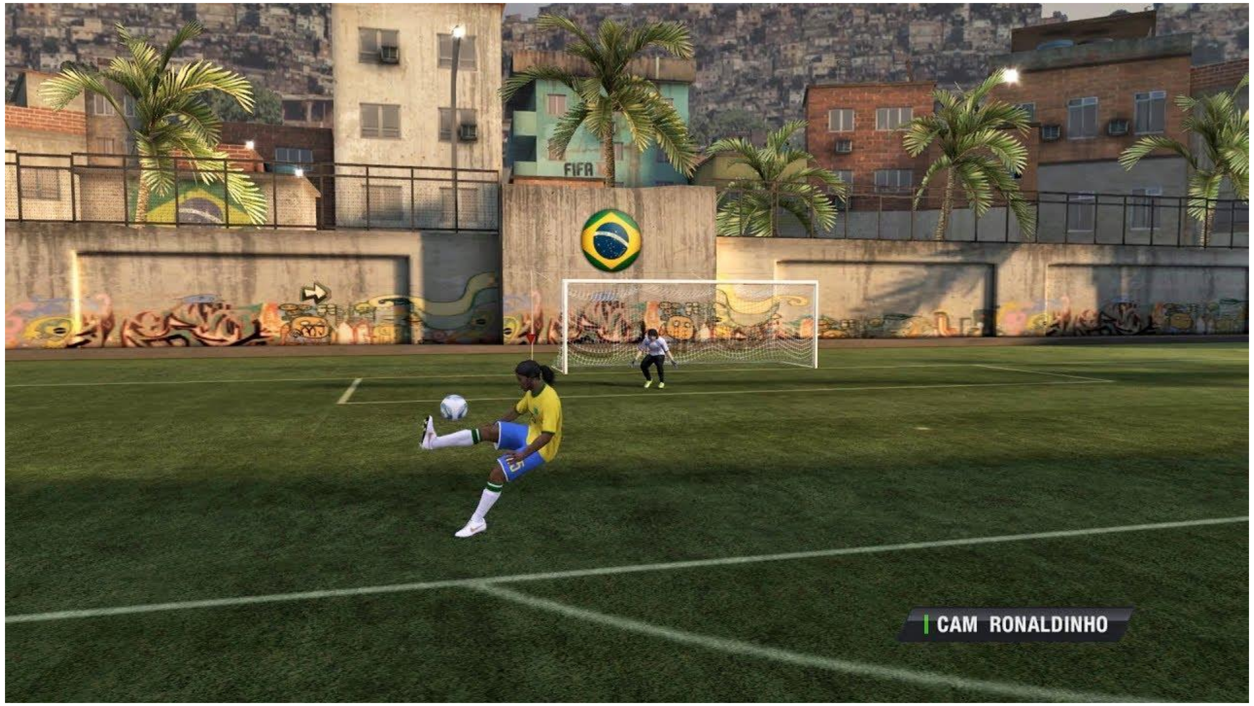
choice: Right if $\hat{s} > 0$, Left otherwise (with probability 1-λ)
randomly chosen between Right and Left (with probability λ)



Part II
Fitting your model to data

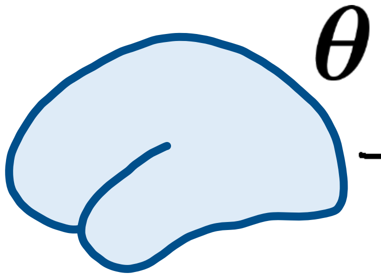
BAMB! Summer School
Lecture I

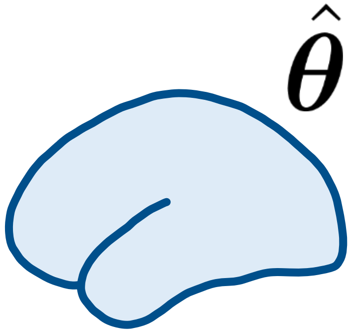
On the importance of simulating your model



Before fitting your model to data, simulate it (with various parameter sets) to **develop intuitions** about its behavior and understand how it is affected by the choice of parameters.

Parameter estimation (*aka* inference)

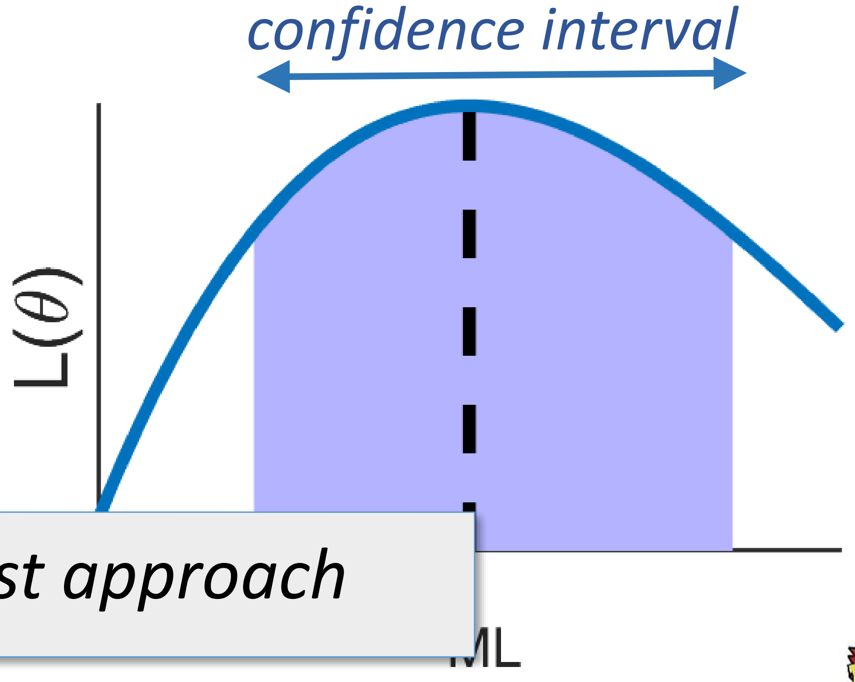
Simulation: $\boldsymbol{x}_t \longrightarrow$  $\longrightarrow \boldsymbol{y}_t$

Inference: $(\boldsymbol{X}, \boldsymbol{Y}) = \{(\boldsymbol{x}_1, \boldsymbol{y}_1), \ldots, (\boldsymbol{x}_n, \boldsymbol{y}_n)\} \longrightarrow$ 

We want the parameters with better match between *model* behavior and *observed* behavior.

-> Optimize measure of *how plausible would the experimental data be if the model were to be true,* i.e. maximize **likelihood**:

$$L(\boldsymbol{\theta}) = p(\boldsymbol{Y} | \boldsymbol{X}, \theta)$$



That's the frequentist approach

Maximum Likelihood Estimation

Computing model likelihood

+

Optimization procedure

Maximum likelihood estimation

$$\hat{\boldsymbol{\theta}}_{ML} = arg\,\max\nolimits_{\boldsymbol{\theta}} L(\boldsymbol{\theta})$$

log-likelihood (LLH)

$$\hat{\boldsymbol{\theta}}_{ML} = arg\,\max\nolimits_{\boldsymbol{\theta}} \log L(\boldsymbol{\theta}) = arg\,\max\nolimits_{\boldsymbol{\theta}} LL(\boldsymbol{\theta})$$

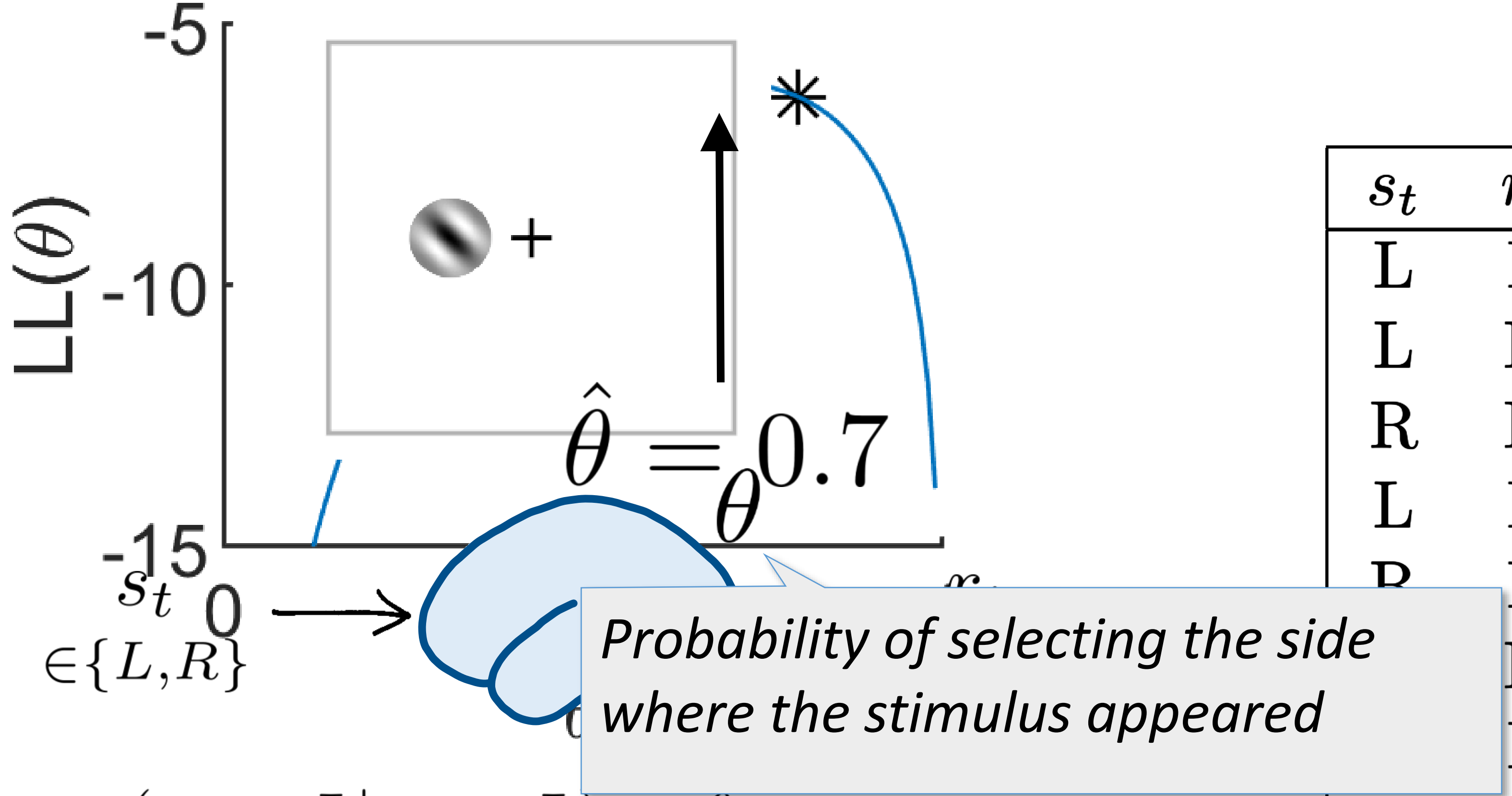
If behavior is *independent* across trials, LLH turns into sum of terms:

$$L(\boldsymbol{\theta}) = p(\mathbf{Y}|\mathbf{X}, \theta) \quad \ldots \quad \boldsymbol{x}_n, \boldsymbol{\theta})$$

The sum is more practical to deal with for analytical and numerical reasons

$$\begin{aligned} LL(\boldsymbol{\theta}) &= \log p(\boldsymbol{y}_1|\boldsymbol{x}_1, \boldsymbol{\theta}) + \ldots + \log p(\boldsymbol{y}_n|\boldsymbol{x}_n, \boldsymbol{\theta}) \\ &= \log f_{\mathcal{M}}(\boldsymbol{y}_1; \boldsymbol{x}_1, \boldsymbol{\theta}) + \ldots + \log f_{\mathcal{M}}(\boldsymbol{y}_n; \boldsymbol{x}_n, \boldsymbol{\theta}) \end{aligned}$$

Toy example: 2 AFC detection task



$$p(r_t = L | s_t = L) = \theta$$

$$p(r_t = R | s_t = L) = 1 - \theta$$

$$p(r_t = L | s_t = R) = 1 - \theta$$

$$p(r_t = R | s_t = R) = \theta$$

$\theta = 0.6$

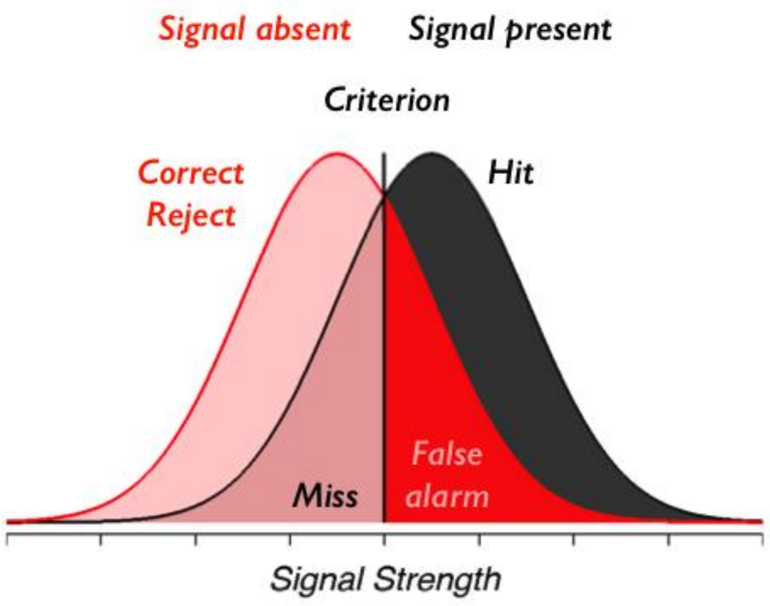
s_t	r_t	$p(r_t = L)$	$p(r_r = R)$
L	L	0.6	0.4
L	R	0.6	0.4
R	R	0.4	0.6
L	L	0.6	0.4
R	L	0.4	0.6
	R	0.4	0.6
	L	0.6	0.4
R	R	0.4	0.6
L	L	0.6	0.4
R	L	0.4	0.6

$$L(\theta{=}0.6) = 0.6 \times 0.4 \ldots \times 0.4 = 0.0018$$

$$LL(\theta{=}0.6) = -6.32$$

BAMB!

Signal detection theory



Generative model

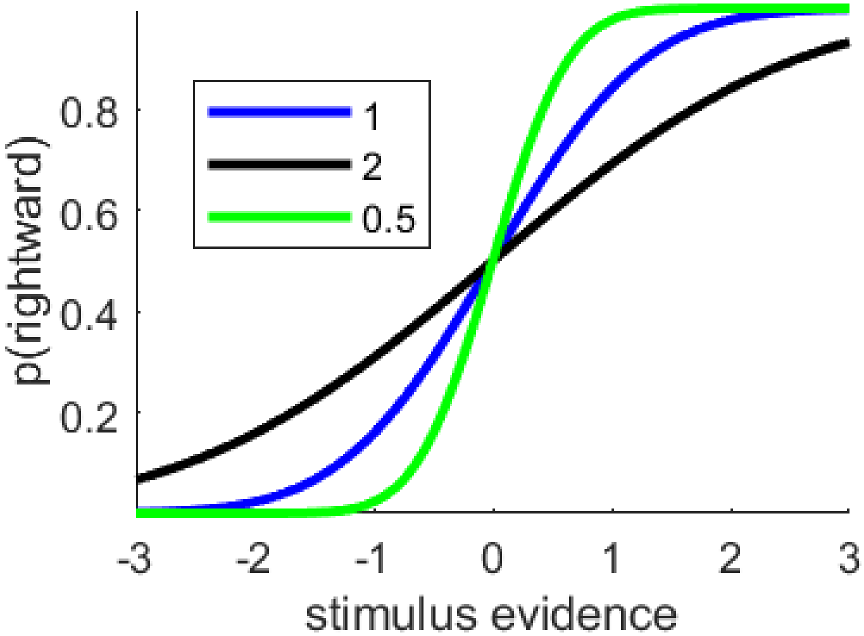
percept $\hat{s} = s + \varepsilon \quad \varepsilon \sim \mathcal{N}(0, \sigma^2)$

choice $r = sign(\hat{s})$

Likelihood

$$p(r = 1|s) = \int_0^{+\infty} p(\hat{s}|s) d\hat{s} = \Phi(s/\sigma)$$

marginalization over latent variable



References

- Palminteri, S., Wyart, V., & Koechlin, E. (2017, June 1). The Importance of Falsification in Computational Cognitive Modeling. *Trends in Cognitive Sciences*, Vol. 21, pp. 425–433. https://doi.org/10.1016/j.tics.2017.03.011
- Wilson, R., & Collins, A. G. E. (n.d.). *Ten simple rules for the computational modeling of behavioral data*. https://doi.org/10.31234/OSF.IO/46MBN
- Nassar, M. R., & Frank, M. J. (2016). Taming the beast: Extracting generalizable knowledge from computational models of cognition. *Current Opinion in Behavioral Sciences*, Vol. 11, pp. 49–54. https://doi.org/10.1016/j.cobeha.2016.04.003
- Bishop, C. M. (2006). Pattern Recognition and Machine Learning, Chapter 3. In *Vasa*. Retrieved from http://medcontent.metapress.com/index/A65RM03P4874243N.pdf