Universitätsklinikum Hamburg-Eppendorf

Model-based fMRI

(cognitive modeling and neuroscience)

SPM Kurs 2011
Expertenrunde

Sebastian Gluth
Department of Systems Neuroscience, UKE

22. 09. 2011

Overview

Part 1: Theory

- Motivation for modeling cognition (and fMRI signals)
- Setting up a cognitive model (reinforcement learning)
- Fitting a model (to behavioral data)
- Model comparison / selection
- Combining modeling results with fMRI

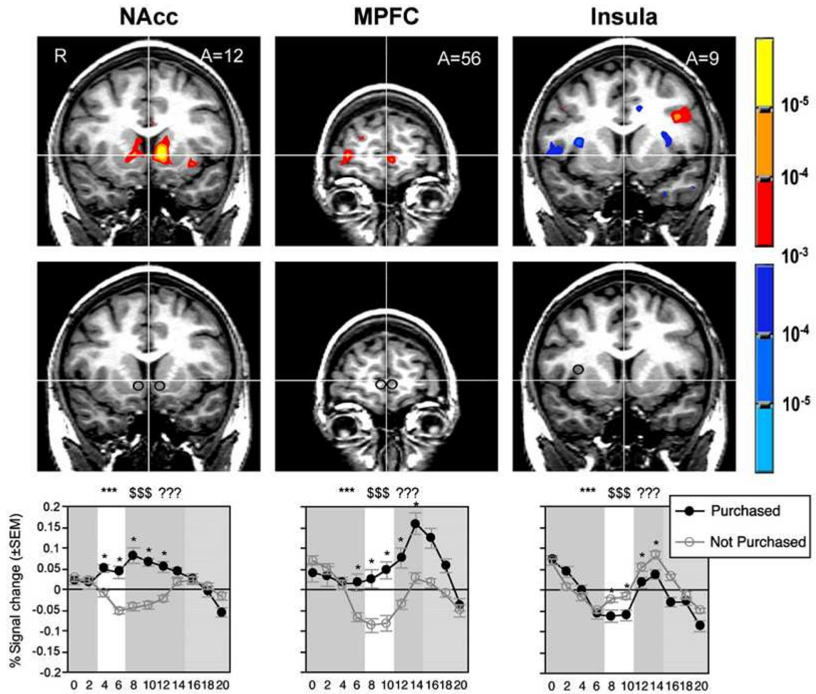
Part 2: Practice

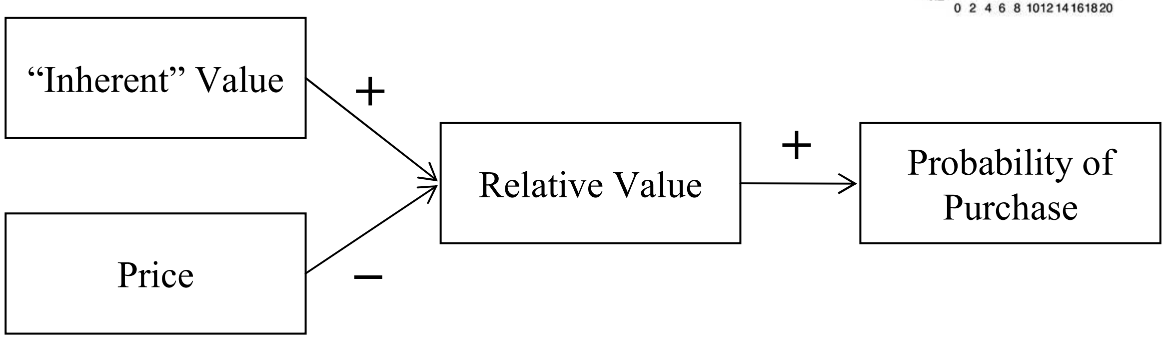
1.1 Motivation for modeling cognition (and fMRI signals)

Motivation

Actually, (almost) every psychological hypothesis or test contains a “model”

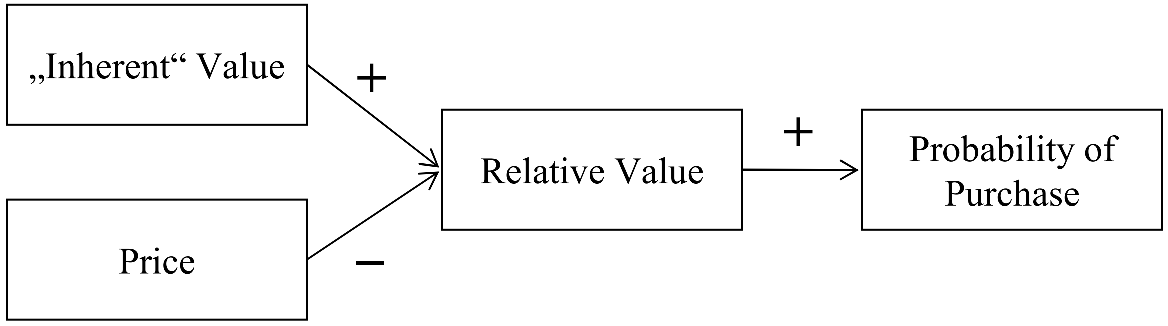
Knutson et al., 2007, Neuron





Motivation

Cognitive modeling tries to shed light on such arrows…



…by making assumption about underlying (cognitive) processes

…practically, by using equations to make testable predictions

→ Better understanding of the “black box” (just like neuroscience)

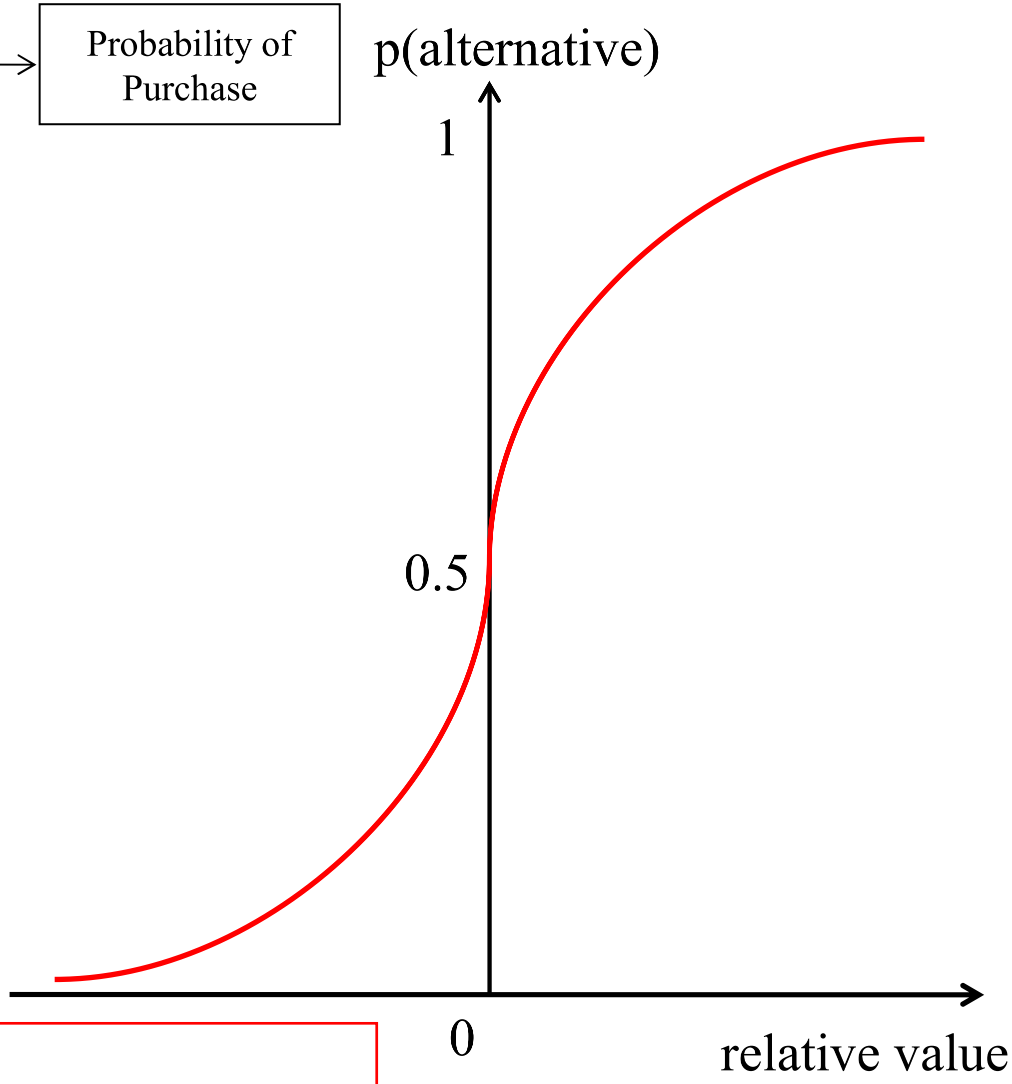
→ Allows trial-by-trial predictions (of behavior and brain signals)

Universitätsklinikum Hamburg-Eppendorf

A simple example

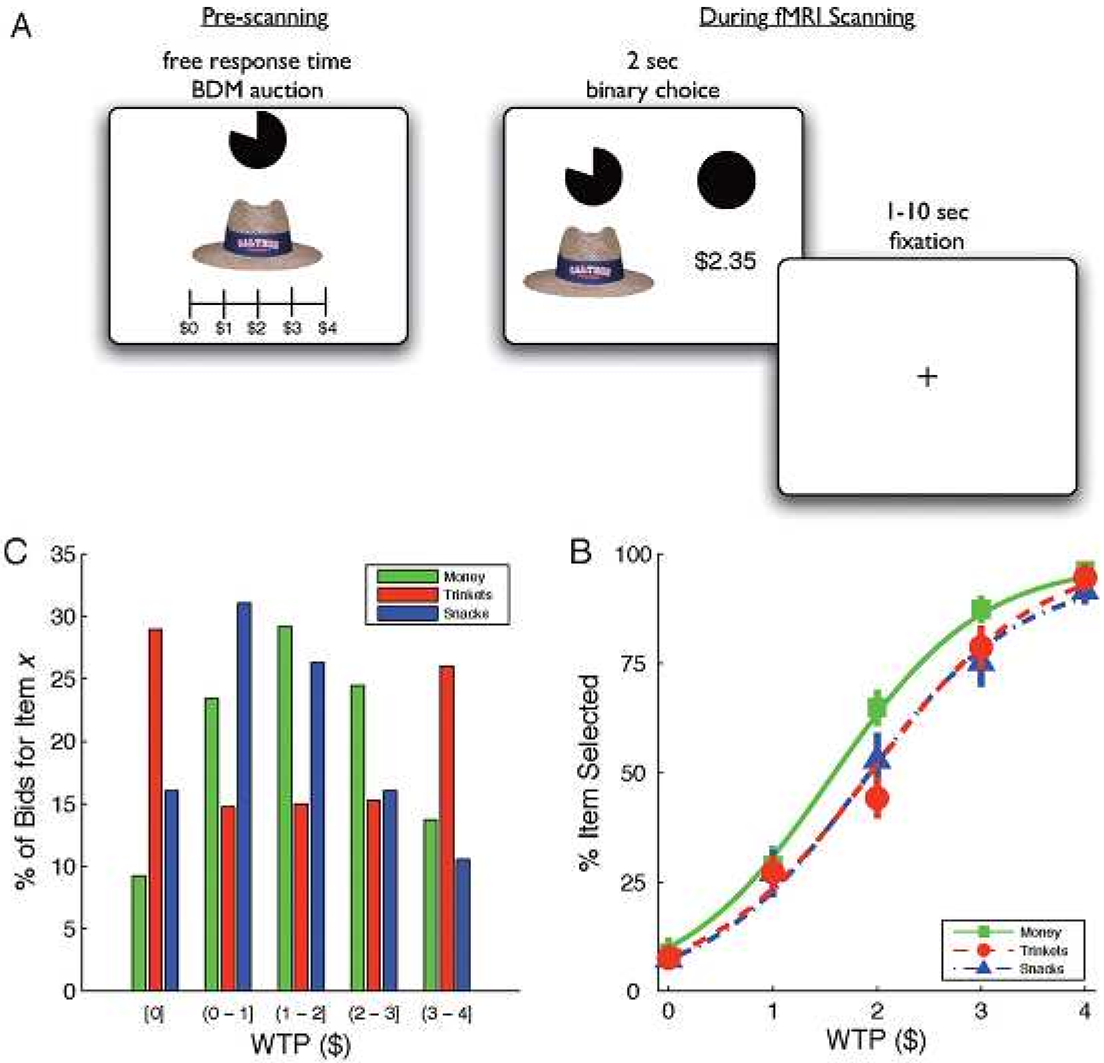
Relative Value + → Probability of Purchase

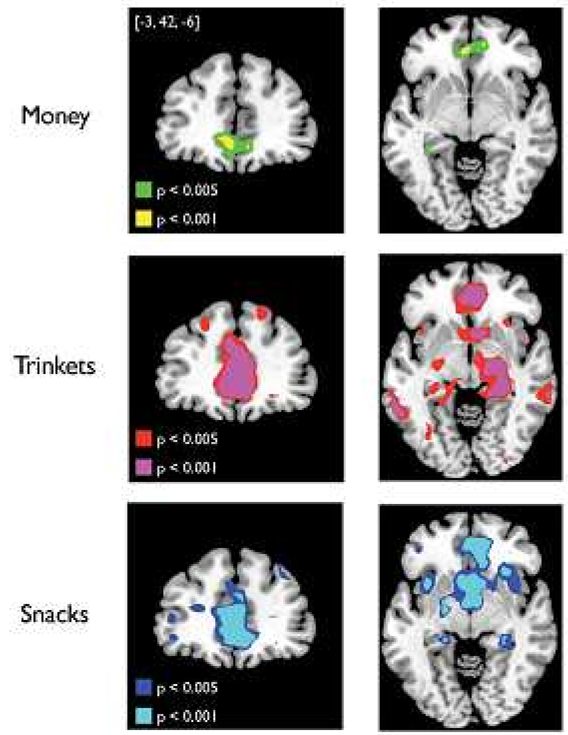
worse ↔ better



$$p(alternative) = \frac{1}{1 + e^{-\gamma \times [Value(alternative) - Value(reference)]}}$$

A simple example

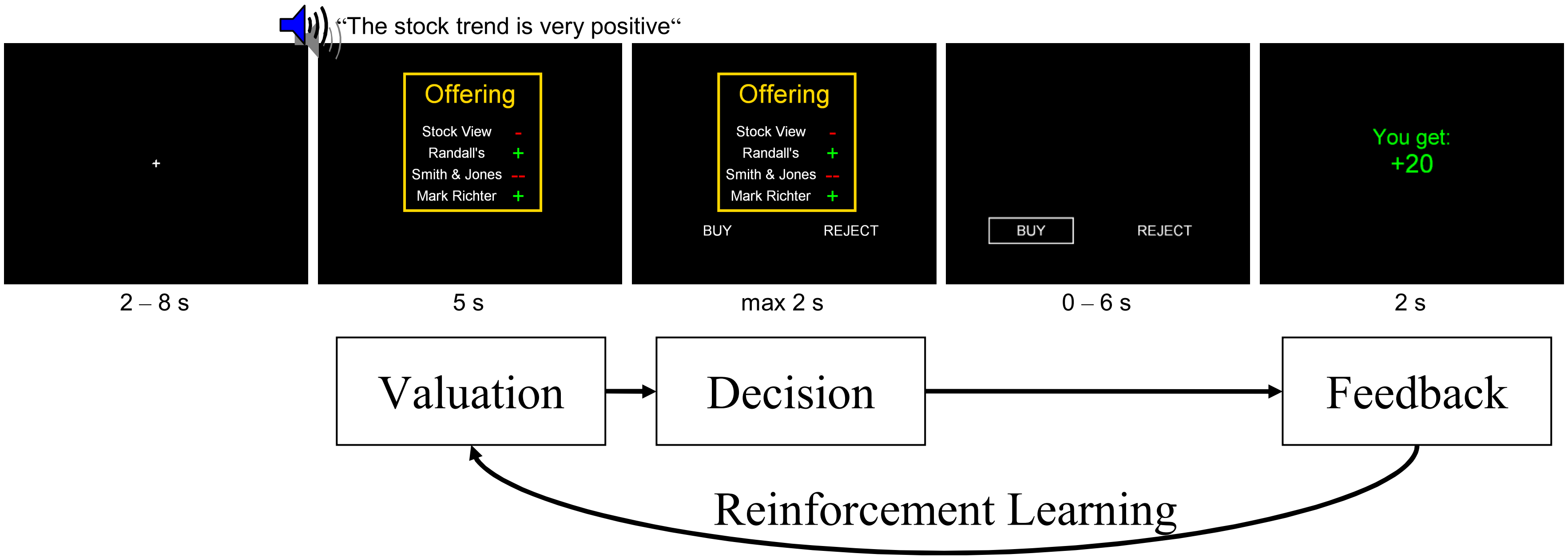


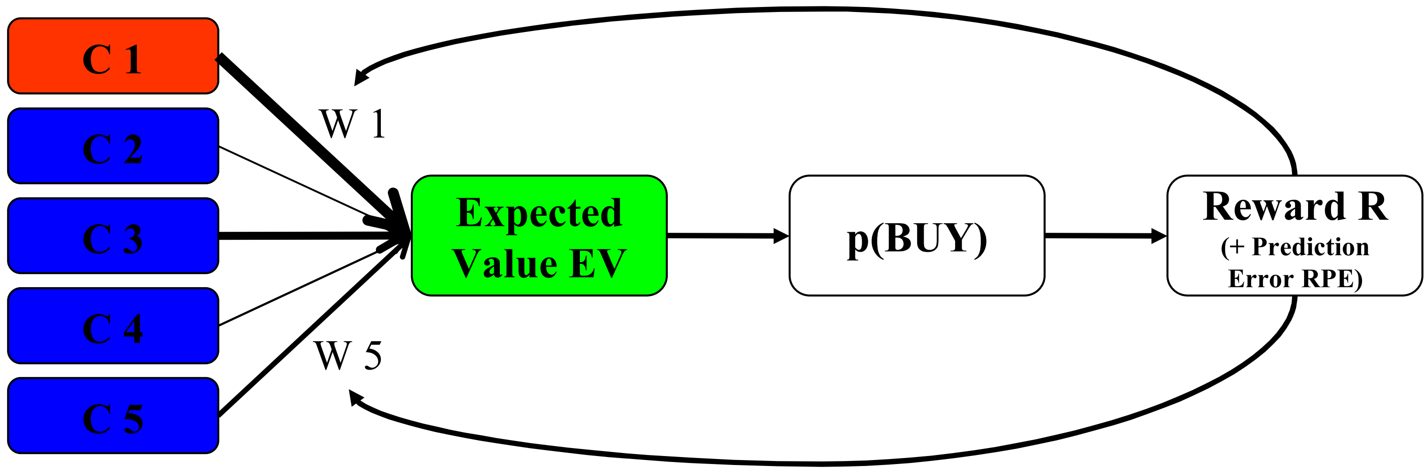


Chib et al., 2009, J. Neurosci.

1.2 Setting up a cognitive model (reinforcement learning)

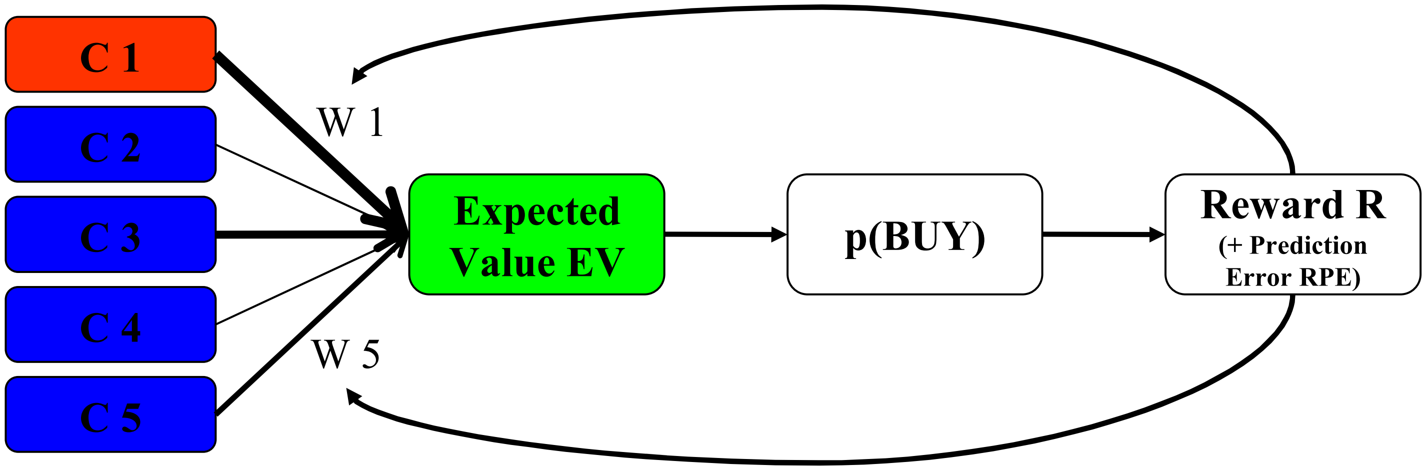
Task and model





Associative linear model, *Adaptive network model*, *Connectionist model* (Gluck & Bower, 1988, JEP; Kelley & Busemeyer, 2008, J. Math. Psych.; Rieskamp, 2006, JEP)

Task and model



Valuation: $EV_t = c_{1,t} \times w_{1,t} + c_{2,t} \times w_{2,t} + c_{3,t} \times w_{3,t} + c_{4,t} \times w_{4,t} + c_{5,t} \times w_{5,t} = \sum_{m=1}^{5} c_{m,t} \times w_{m,t}$

Decision: $p(BUY)_t = \frac{1}{1+e^{-\gamma \times EV_t}}$

"Surprise": $RPE_t = R_t - EV_t \times I_t$

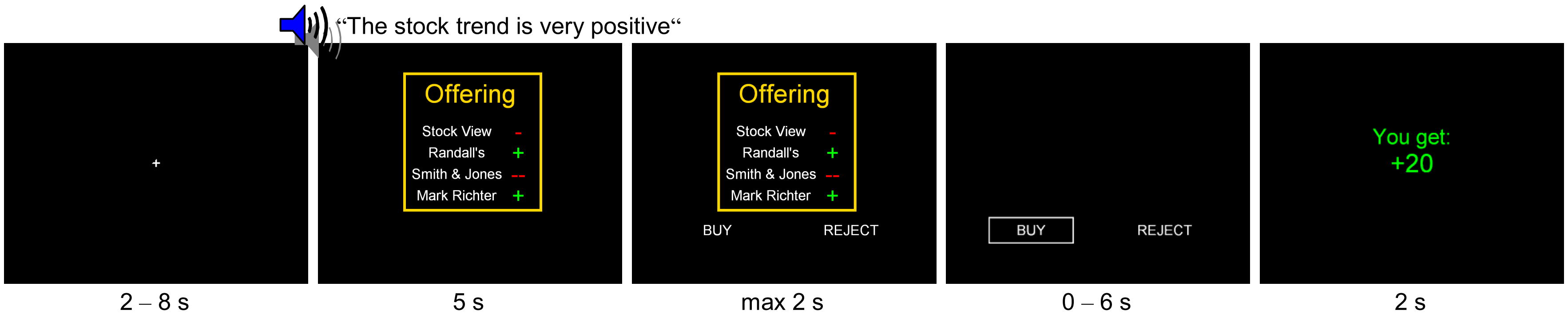
Learning: $w_{m,t+1} = w_{m,t} + \alpha \times RPE_t \times c_{m,t}$

free parameters:

γ → stochasticity parameter
(or: inverse temperature)

α → learning rate

A single trial



Cues:

$c1 = +2$

$c2 = -1$

$c3 = +1$

$c4 = -2$

$c5 = +1$

Weights:

$w1 = 7.0$

$w2 = 0.5$

$w3 = 5.5$

$w4 = 8.0$

$w5 = 0.0$

Valuation:

$\mathbf{EV} = c1*w1 + c2*w2 + c3*w3 + c4*w4 + c5*w5 = 2*7-1*0.5+1*5.5-2*8+1*0 = \mathbf{3}$

Decision:

$p(BUY) = 1/(1+\exp(-\gamma*\mathbf{EV})) = 1/(1+\exp(-\mathbf{0.1}*\mathbf{3})) = .57444$

“Surprise”:

$\mathbf{RPE} = R - \mathbf{EV} = 20 - \mathbf{3} = \mathbf{17}$

Learning:

$w1_{t+1} = w1_t + \alpha*\mathbf{RPE}*c1 = 7.0+\mathbf{0.1}*\mathbf{17}*2 = \quad 10.4$

$w2_{t+1} = w2_t + \alpha*\mathbf{RPE}*c2 = 0.5+\mathbf{0.1}*\mathbf{17}*-1 = \quad -1.2$

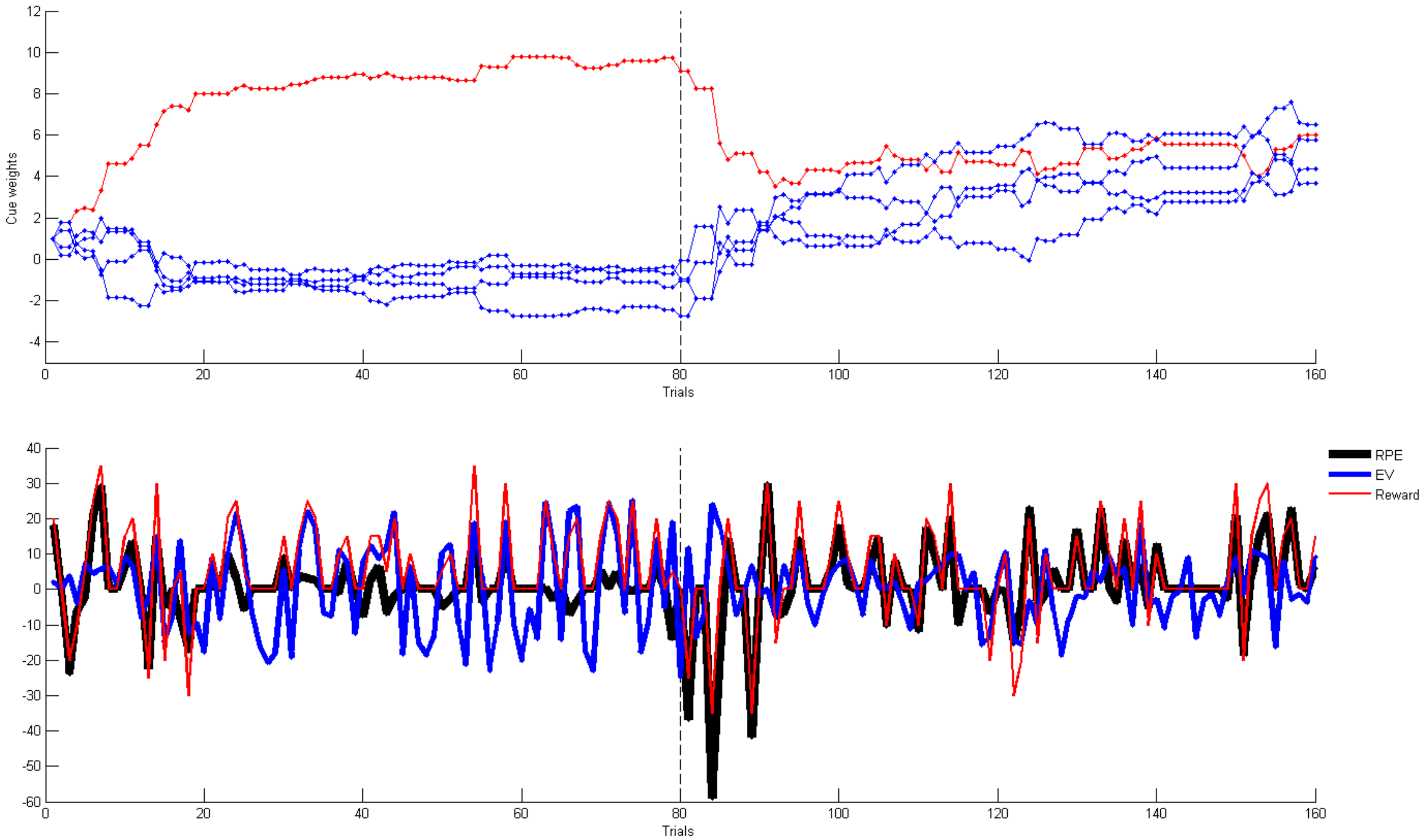
$w3_{t+1} = w3_t + \alpha*\mathbf{RPE}*c3 = 5.5+\mathbf{0.1}*\mathbf{17}*1 = \quad 7.2$

$w4_{t+1} = w4_t + \alpha*\mathbf{RPE}*c4 = 8.0+\mathbf{0.1}*\mathbf{17}*-2 = \quad 4.6$

$w5_{t+1} = w5_t + \alpha*\mathbf{RPE}*c5 = 0.0+\mathbf{0.1}*\mathbf{17}*1 = \quad 1.7$

new weights!

A single subject



1.3 Fitting a model (to behavioral data)

Model fitting

Question: How good does our model describe the data?

Problem: We need a link between model predictions and observable behavior

Not directly observable:

- Expected value
- Reward prediction error
- Cue weights
- Probability to buy / to reject

Observable:

→ Response in each trial (bought or rejected)

Solution: Link between p(BUY) and responses

[good: high p(BUY) when bought, low p(BUY) when rejected]

[bad: low p(BUY) when bought, high p(BUY) when rejected]

Universitätsklinikum Hamburg-Eppendorf

Model fitting

Difference between model and data:

$$\Delta = \left| p(BUY) - I \right|$$

I = 1, when bought
I = 0, when rejected

Examples:

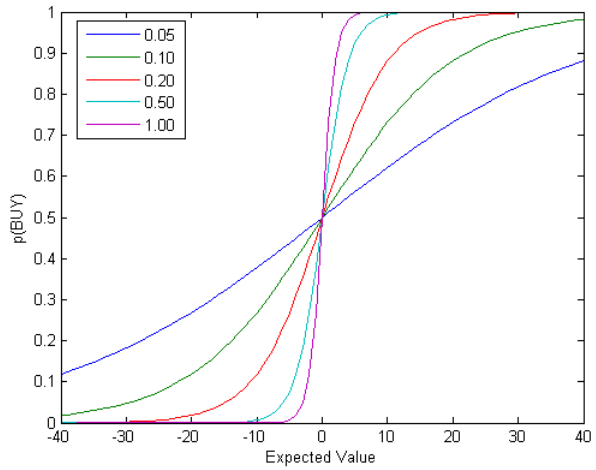
$p(BUY) = 0.9, I = 1$ → $\Delta = 0.1$ (good)
$p(BUY) = 0.1, I = 1$ → $\Delta = 0.9$ (bad)
$p(BUY) = 0.8, I = 0$ → $\Delta = 0.8$ (bad)
$p(BUY) = 0.2, I = 0$ → $\Delta = 0.2$ (good)

Goal: minimize Δ by using the best parameter values

Parameter values? → specific values for free parameters

Free parameters? → learning rate (α) and stochasticity parameter (γ)

$w1_{t+1} = w1_t + \alpha1 * RPE * c1 = 7 + \mathbf{0.1} * 17 * 2 = 7 + \mathbf{3.4} = \mathbf{10.4}$
$w1_{t+1} = w1_t + \alpha2 * RPE * c1 = 7 + \mathbf{0.5} * 17 * 2 = 7 + \mathbf{17} = \mathbf{24}$



Free parameters allow to account for inter-individual differences!!!
(also: differences between conditions)

Parameter estimation

Ways to find good parameters:

1. Selecting a single value based on:
   - Previously published parameter values
   - "Biological plausibility" (O'Doherty et al., 2003, Neuron)

2. Iterative search for best parameter values:
   - Grid search
   - Automatic optimization (e.g., *fminsearch* function in Matlab)

state of the art!

A cost/discrepancy function is required for the 2nd method:

- Least squares (RMSD = root mean square deviation)
- Maximum likelihood estimation

Maximum Likelihood Estimation

$\Delta = \left| p(BUY) - I \right|$

Δ = ”inverse of the likelihood of the data given the model”

$Likelihood = 1 - \Delta = 1 - \left| p(BUY) - I \right|$

→ We want to maximize the likelihood, so we want to minimize Δ

Better:

maximize the log-likelihood (logarithm of the likelihood)

$Loglikelihood = \ln\left(1 - \left| p(BUY) - I \right|\right)$

Reasons:

logarithm compresses likelihoods within reasonable ranges

with more observations, models are easier to distinguish

Maximum Likelihood Estimation

Likelihoods can get extremely small, particularly with many observations:
let‘s say, we predict choices in 200 trials with a probability of 60%

The total likelihood will be:

$0.6^{200} = 4.27*10^{-45}$ =0.000000000000000000000…427

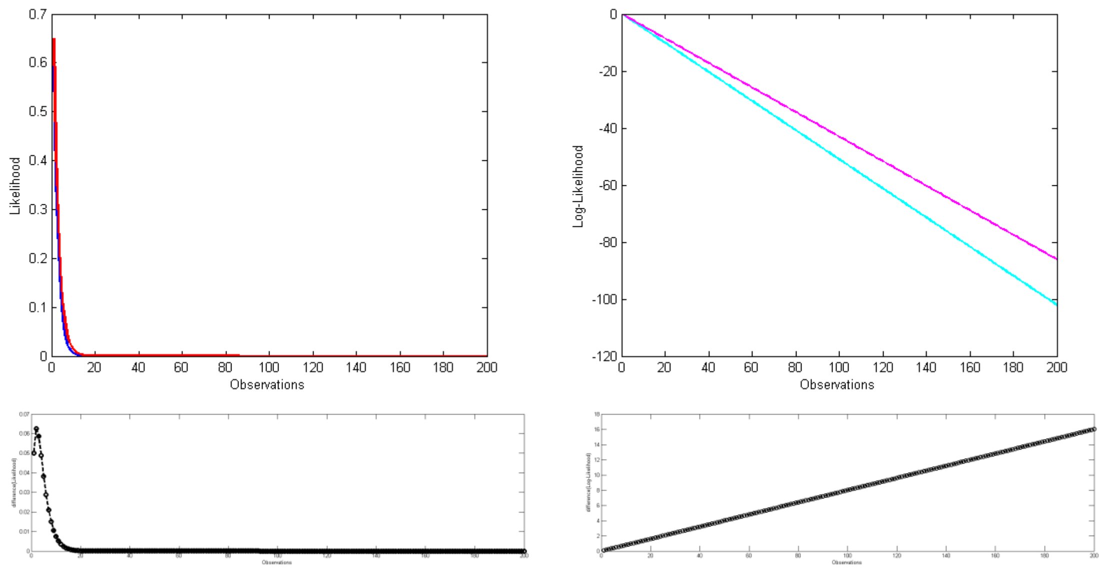
In contrast, log-likelihoods are added: $\prod_{x=1}^{X} f(x) = \sum_{x=1}^{X} \ln(f(x))$

→ $\log(0.6^{200}) = \log(0.6)*200 = -102.17$

Two advantages:

- reasonable values (Matlab cannot really work with values like $4.27*10^{-45}$)
- differences between models, that truely differ, increase with the number of observations

Blue/Cyan: p = 60%
Red/Magenta: p = 65%

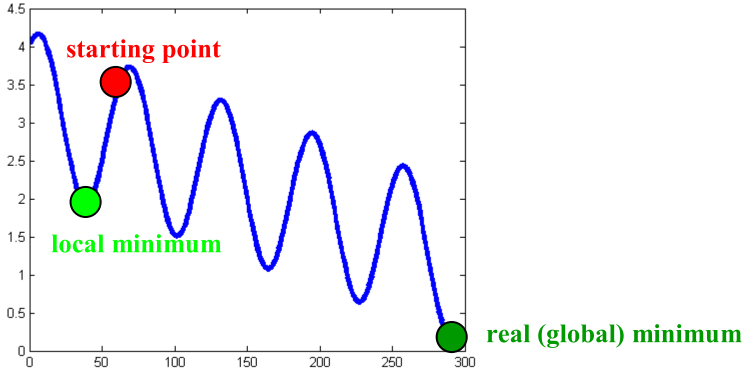


Maximum Likelihood Estimation

Automatic optimization via **fminsearch** in Matlab

- Uses the non-linear SIMPLEX algorithm (Nelder & Mead, 1965)
- **fminsearch** means "**search** for **min**imum of a **f**unction"
- But we want to **max**imize the (log-)likelihood
- → minimize the deviance: $G^2 = -2 \times Loglikelihood = -2 \times \sum_{t=1}^{160} \ln\left(1 - |\, p\left(BUY\right)_t - I_t \,|\right)$

- Danger of getting stuck in local minima:



Solutions: use different starting values for parameters

grid search to find best starting values

Model fitting

Granularity:

1. Fitting group data (one parameter set for all subjects)
2. Fitting individual data (estimating parameter sets for each subject)
3. Fitting data of separate conditions (e.g., to show that learning rate is different)

→ I recommend individual parameter estimation

Jan Gläscher: “Fixed effects” model-based fMRI might yield more stable results

1.4 Model comparison / selection

How good is a model?

$$G^2 = -2 \times Loglikelihood = -2 \times \sum_{t=1}^{160} \ln\left(1 - |\, p(BUY)_t - I_t \,|\right)$$

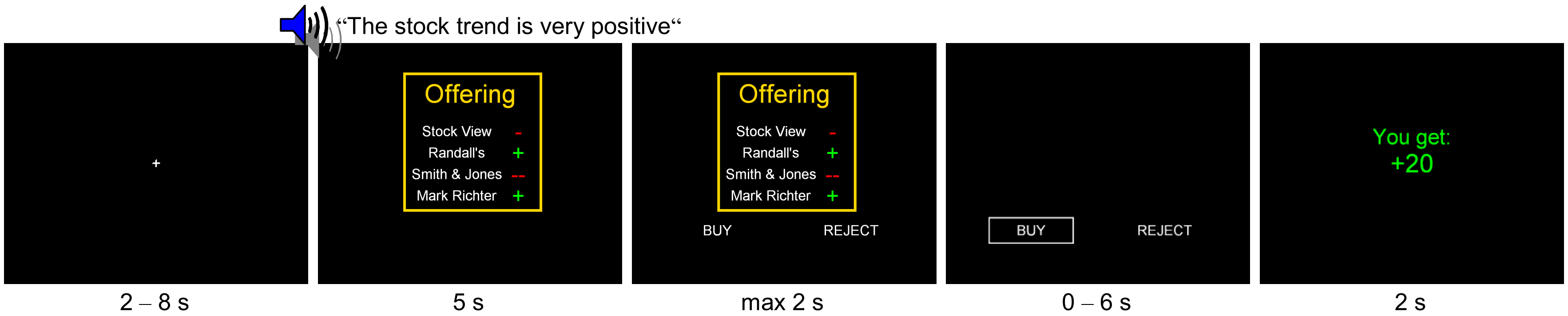
→ deviance and log-likelihood are values that don't tell us much (at first)

(e.g., the deviance is 131.70, the log-likelihood is -65.85 → ???)

Usually, models are compared against each other to say which one is best ("relative model goodness")

→ different models imply different views on cognitive / neural processes

Our example



So far, we have assumed one learning rate for all cues $w_{m,t+1} = w_{m,t} + \alpha \times RPE_t \times c_{m,t}$

What about two independent learning rates for auditory and visual cues?
→ +1 parameter
→ nested models (the more complex model is a *generalization* of the easier model)

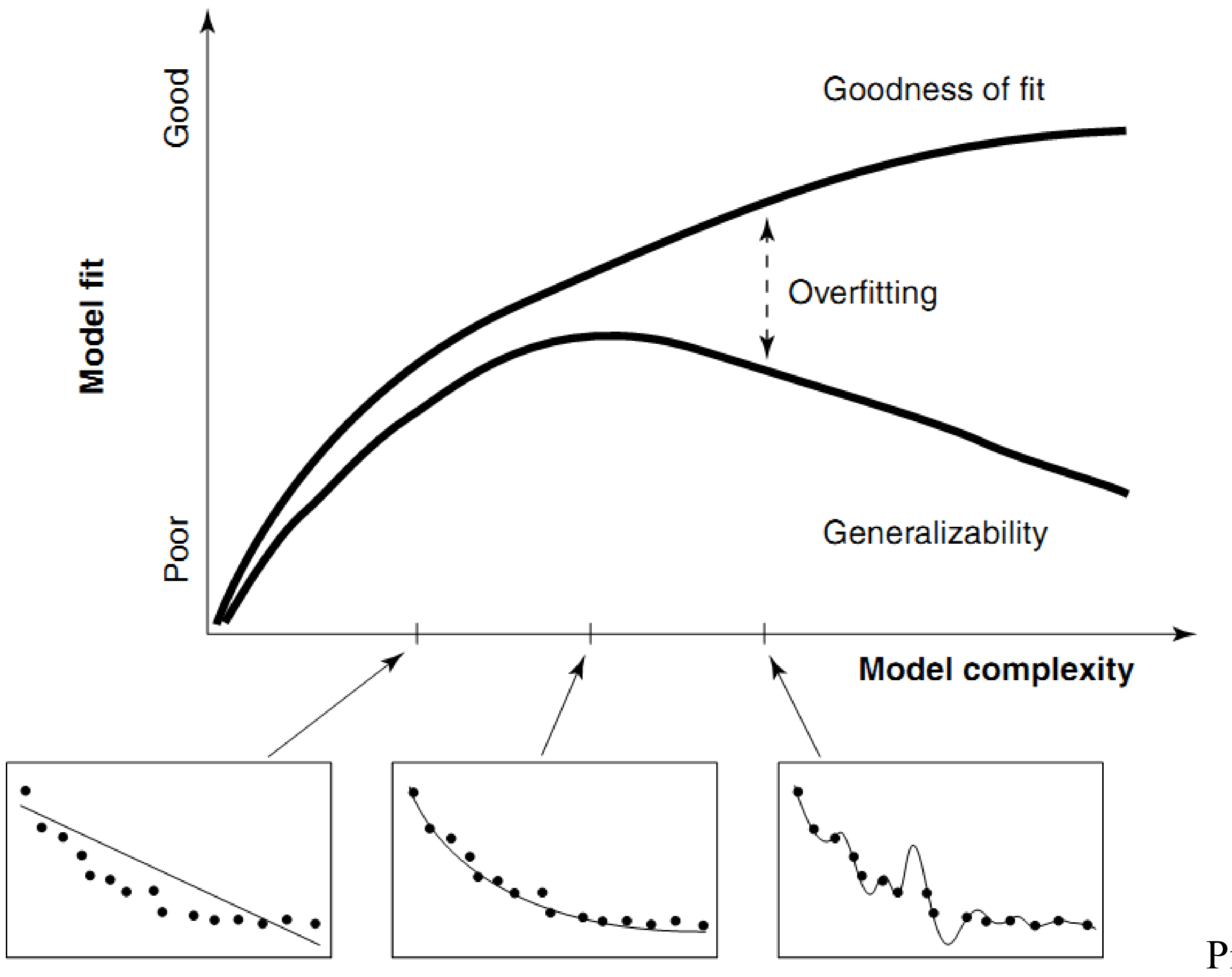
For nested models: $fit(complex) \geq fit(specific)$

→ Question of **significance** of better model fit

Log-likelihood and deviance can be used to make statistical inferences!!!

Model comparison

Model fit vs. Generalizability



Pitt & Myung, 2002, TICS

Model comparison

Comparison of nested models:

→ Chi²-Test of the difference of the deviance $G^2(complex) - G^2(specific)$

(the random improvement of the deviance by introducing *n* additional parameters is approximately Chi²-distributed with *df* = *n*)

Comparison of non-nested models (more common):

Akaike Information Criterion (AIC) – number of free parameters (*k*)

$$AIC = -2 \times LL + 2k$$

Bayesian Information Criterion (BIC) – number of free parameters & trials (*N*)

$$BIC = -2 \times LL + k \times \ln(N)$$

Minimum Description Length (MDL) – number of free parameters & trials

+ functional form of the model

Also: Bayes Factor, Free Energy

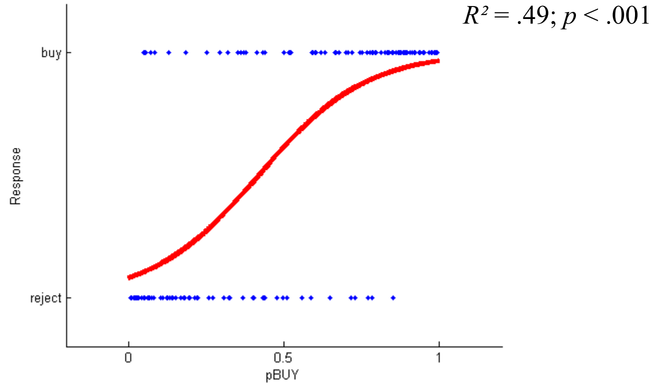
But how good is the model really?

Model comparison and parameter optimization does not guarantee an overall good model fit – what can we say about models in absolute terms?

1. average prediction accuracy
   (should be higher than 50%)

$$\frac{\sum_{t=1}^{n} 1 - \left| p(BUY)_t - I_t \right|}{n}$$

2. (logistic) regression of predictions against data:



3. comparison with baseline model:
   - baseline = chance level
   - for two choice options, „a priori“ chance level = 50%

$$G^2(baseline) = -2 \times \ln(0.5) \times n = 221.81$$

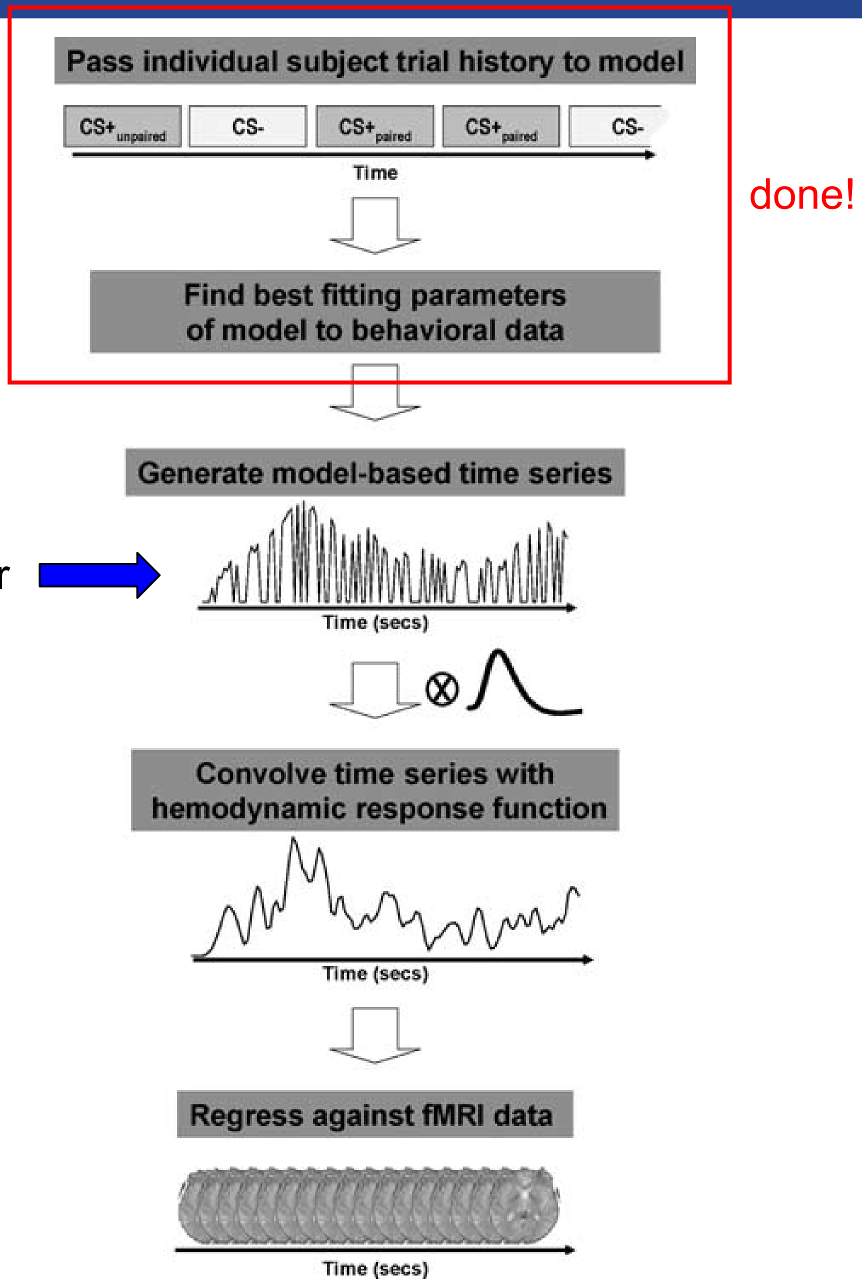
   - better: chance level → choice frequencies (if bought in 75%, chance level should be 75%)
   - allows you to say, that the model performs at least better than chance level

4. correlation of parameter values with behavior (e.g., learning rate with accuracy)

1.5 Combining modeling results with fMRI

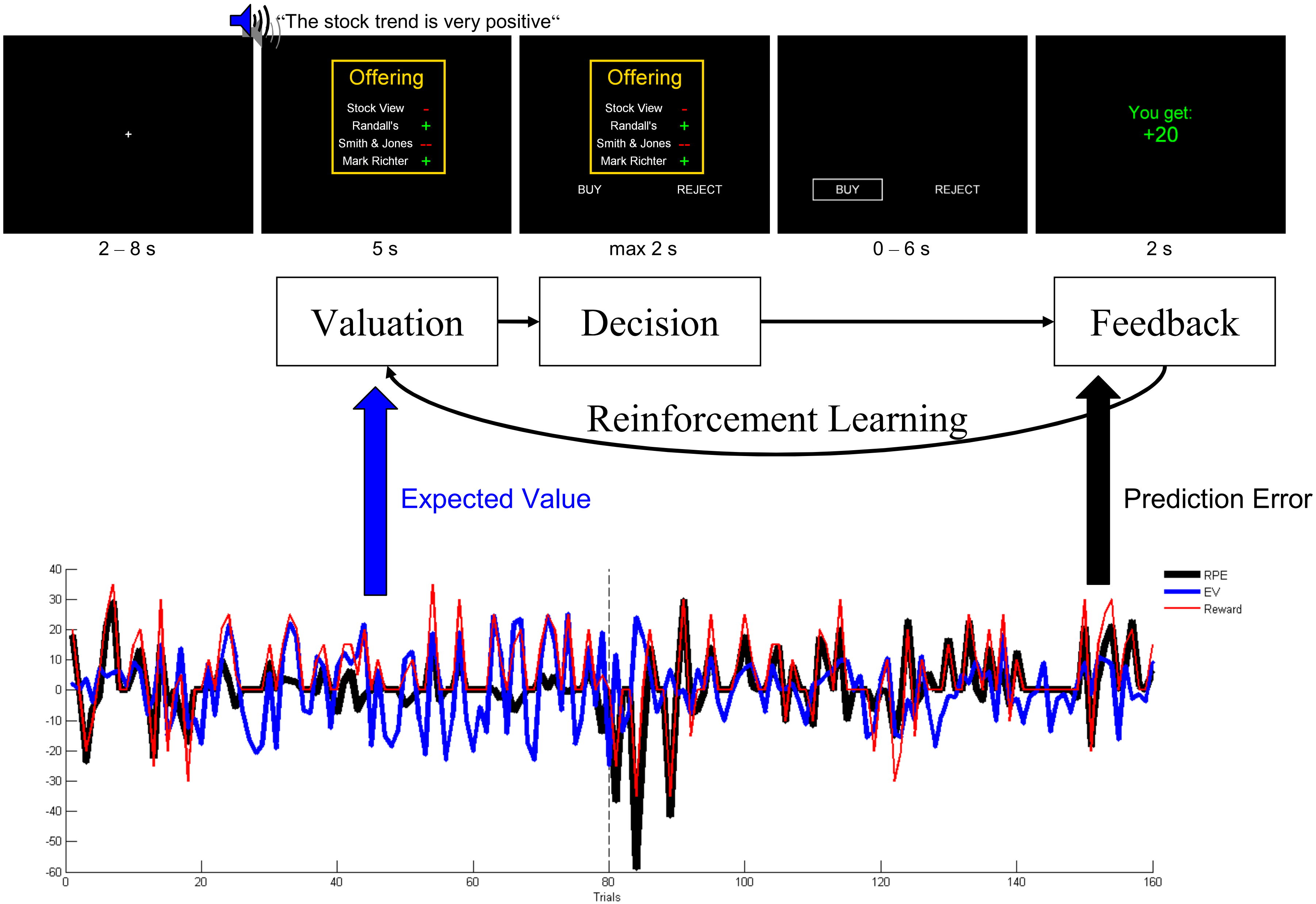
Implementation in SPM

O‘Doherty et al., 2007, Ann. N. Y. Acad. Sci.:



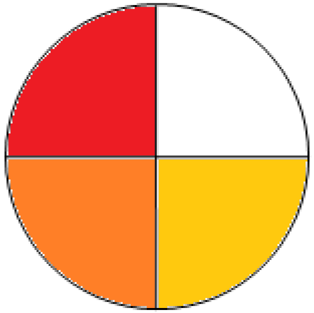
implementation in SPM: parametric modulator of an existing onset regressor

Our onsets and parametric modulators



Implementation in SPM

Be careful:
If there are multiple parametric modulators for one onset regressor, SPM will automatically orthogonalize from left to right



Solutions:
1.Switch order of modulators (and just look at the last modulator)
→ Problem: too many first-level analyses

2.Each modulator gets its own onset regressor
→ Problem: Effects for onsets cannot be estimated anymore

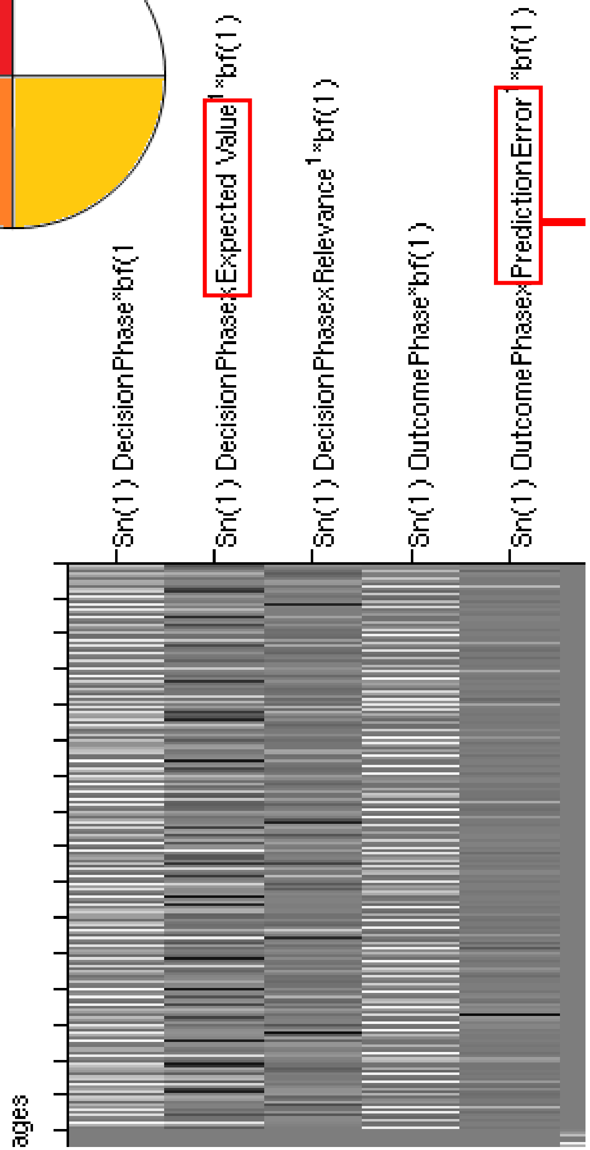
3.Uncomment orthogonalization in SPM:

spm_get_ons (line 228):

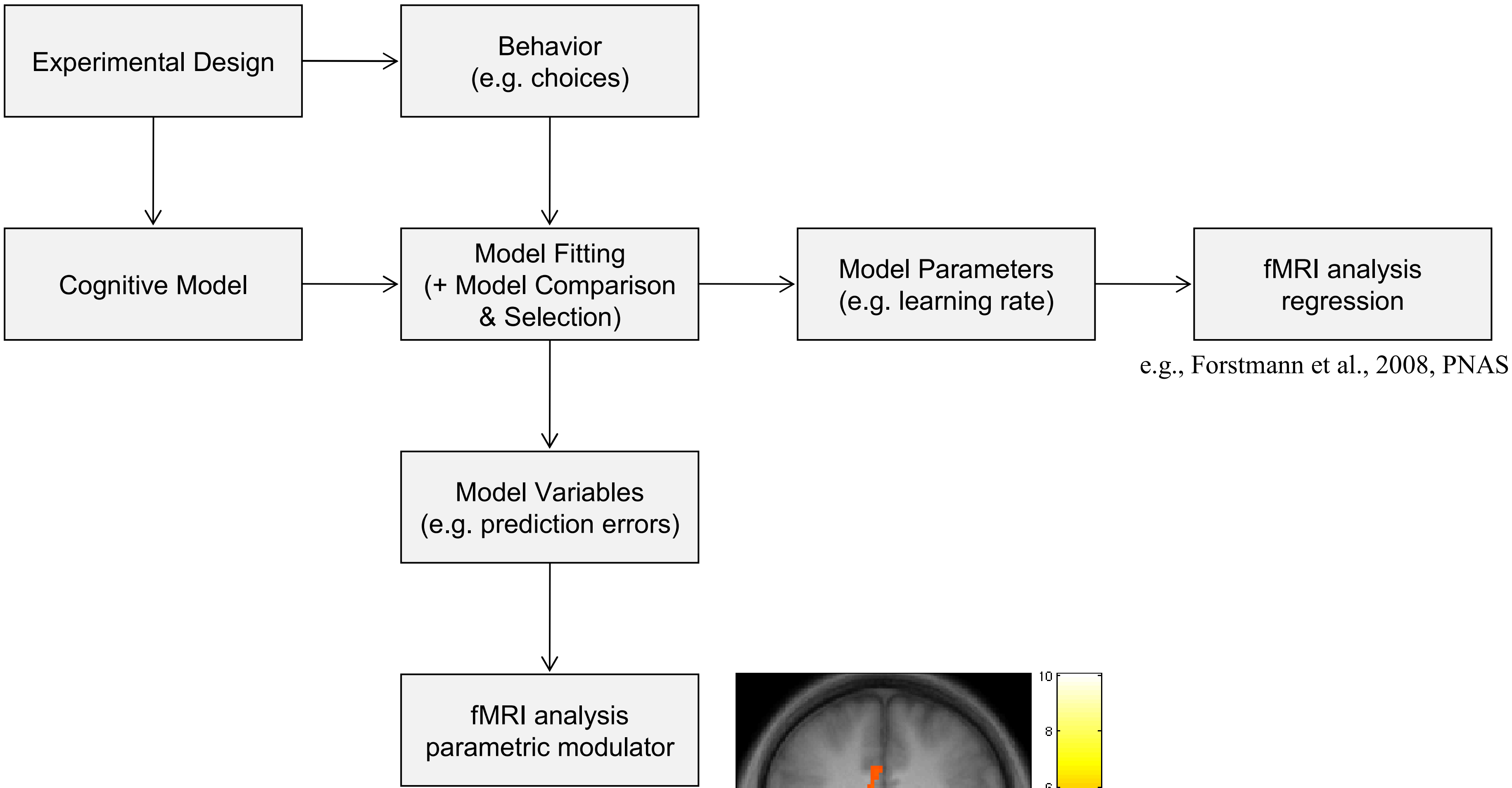
```
226      % orthogonalize inputs
227      %----------------------------------------------------------
228 -    u      = spm_orth(u);
```

spm_fMRI_design (lines 277 – 279):

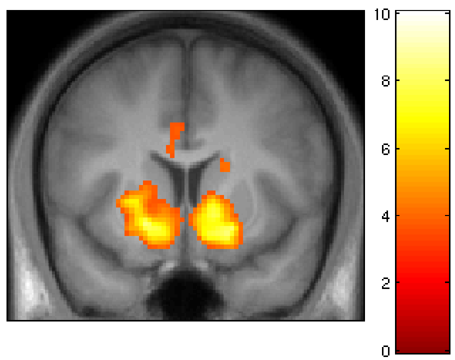
```
275          % and orthogonalise (within trial type)
276          %----------------------------------------------------------
277 -        for i = 1:length(Fc)
278 -            X(:,Fc(i).i) = spm_orth(X(:,Fc(i).i));
279 -        end
```



Summary



Experimental Design
Behavior
(e.g. choices)
Cognitive Model
Model Fitting
(+ Model Comparison
& Selection)
Model Parameters
(e.g. learning rate)
fMRI analysis
regression
e.g., Forstmann et al., 2008, PNAS
Model Variables
(e.g. prediction errors)
fMRI analysis
parametric modulator



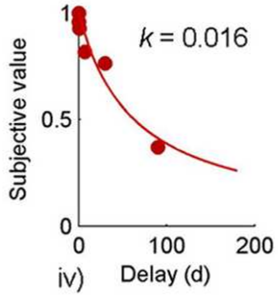
10
8
6
4
2
0

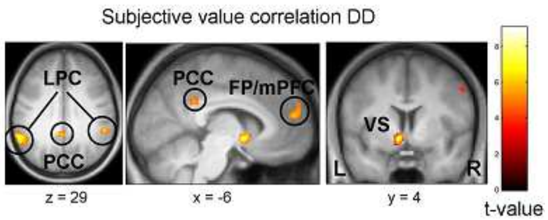
Modeling other stuff

There's more to model than reinforcement learning…

…decision making (e.g., temporal discounting)

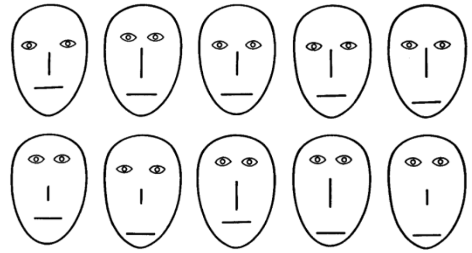
$$SV = \frac{1}{1 + kD}$$



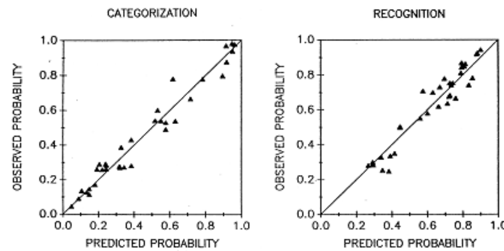


Peters & Büchel, 2009, J. Neurosci.

…classification & recognition memory

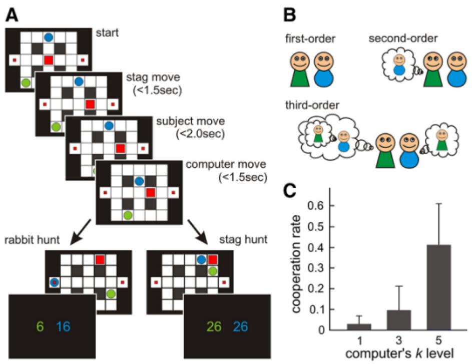


$$d_{ij} = c[\sum_m w_m |x_{im} - x_{jm}|^2]^{1/2}$$



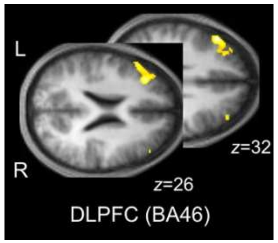
Nosofski, 1991, JEP

…social cognition (game theory)



$$p(k_{com}(T)|y, k_{sub}) \propto p(y(1,\ldots,T)|k_{sub}(1,\ldots,T), k_{com})\,p(k_{com})$$

$$= \prod_{t=1}^{T-1} \kappa^{T-1} p(s_{t+1}|s_t, k_{sub}(t), k_{com})$$



Yoshida et al., 2010, J. Neurosci.

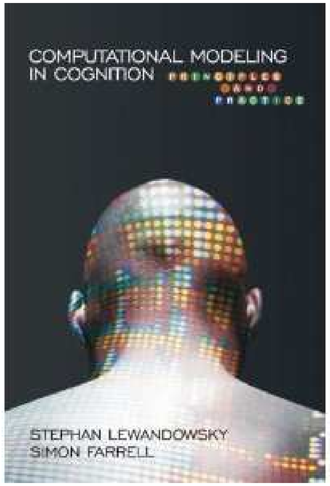
…and more:

- working memory
- perceptual decision making
- multisensory integration
- language
- eye movements
- […]

Literature

Textbook on cognitive modeling (highly recommended!):

Lewandowsky, S., & Farrell, S. (2011). *Computational Modeling in Cognition: Principles and Practice* (Thousand Oaks, CA: Sage).



Model-based fMRI:

O'Doherty, J.P., Hampton, A., & Kim, H. (2007). Model-based fMRI and its application to reward learning and decision-making. *Annals of the New York Academy of Sciences, 1104*, 35–53.

Model fit / model comparison / „what are good and bad models":

Pitt, M.A., & Myung, I.J. (2002). When a good fit can be bad. *Trends in Cognitive Sciences, 6*, 421–425.

Roberts, S., & Pashler, H. (2000). How persuasive is a good fit? A comment on theory testing. *Psychological Review, 107*, 358–367.