CMMC 2018

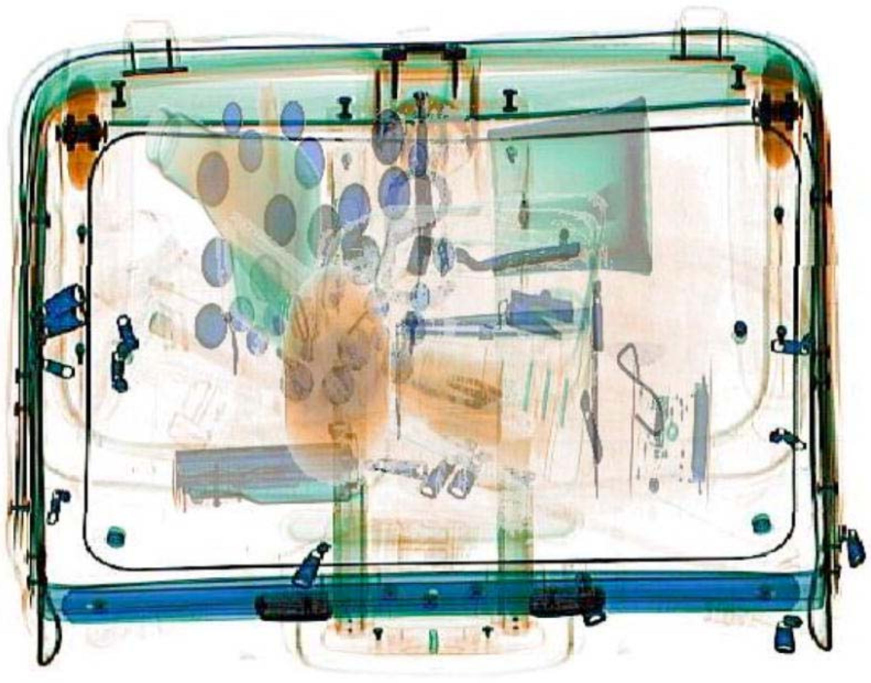
Perceptual decisions tournament

Casimir Ludwig
c.ludwig@bristol.ac.uk

Perceptual decisions tournament

- Introduction to perceptual decision-making (~ 1.5 hrs)
- Tournament introduction (~ 1 hr)
- Lunch
- Model implementation (1.5 hrs)
- Model evaluation

Perceptual decision-making?



P(state of the world | sensory data)

$$P(S|x) = \frac{P(x|S)P(S)}{P(x)}$$

Hallmark characteristics

- Noisy sensory information (external noise)
- Processed by a noisy brain (internal noise)
- Typically: sampling or integrating across space and time (on a variety of spatial and temporal scales; e.g. neuronal to attentional)
- Often: combine sensory information with stored knowledge (i.e. prior beliefs)

Inference problem: what real-world event gave rise to these noisy sensory data?

Models of (perceptual) decision-making

2-AFC:

$$\frac{P(S_1|x)}{P(S_2|x)} = \frac{P(x|S_1)}{P(x|S_2)}\frac{P(S_1)}{P(S_2)}$$

When both options are equally likely:

$$\frac{P(S_1|x)}{P(S_2|x)} = \frac{P(x|S_1)}{P(x|S_2)}$$

Models of (perceptual) decision-making

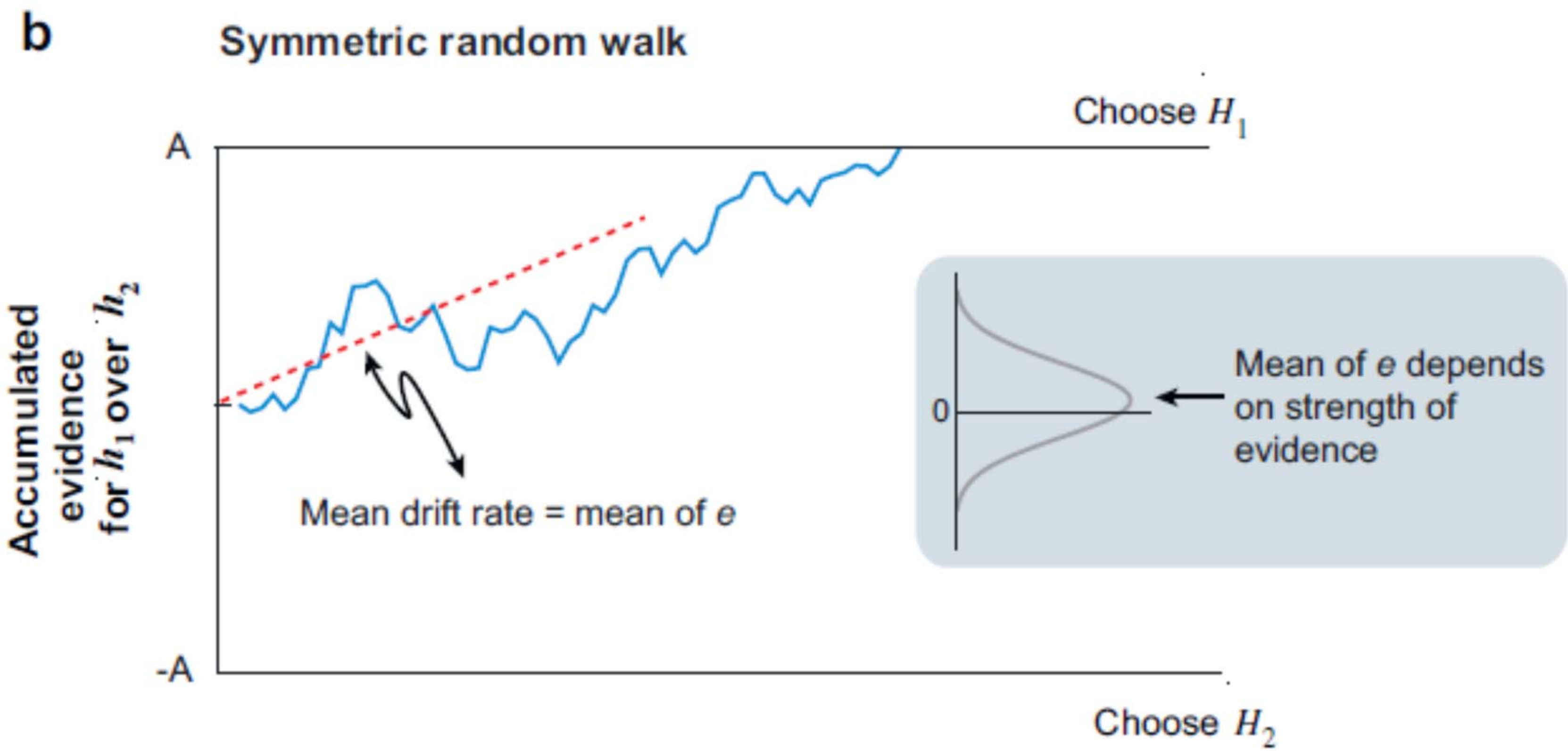
When the evidence comes in over time

$$\frac{P(S_1|\mathbf{x})}{P(S_2|\mathbf{x})} = \frac{\prod_{t=1}^{T} P(x_t|S_1)}{\prod_{t=1}^{T} P(x_t|S_2)} = \prod_{t=1}^{T} \frac{P(x_t|S_1)}{P(x_t|S_2)} = \mathrm{LR}$$

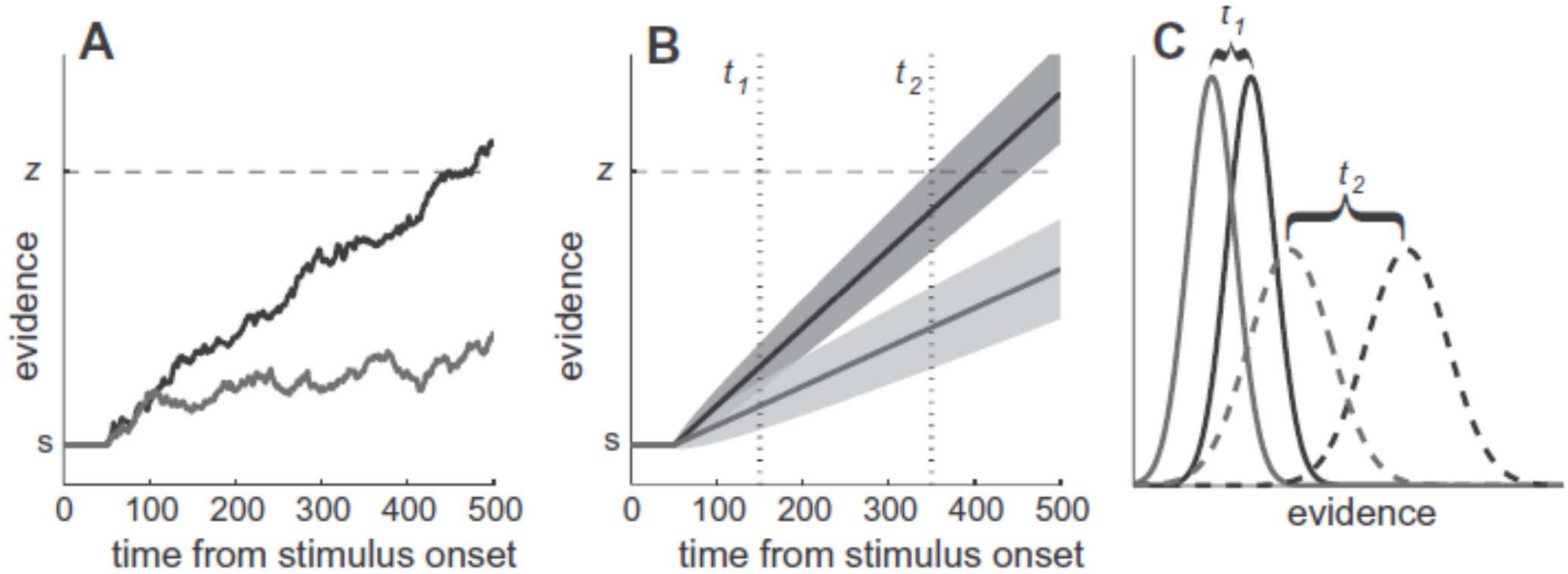
$$\mathrm{LLR}$$
$$= \sum_{t=1}^{T} \ln\left[\frac{P(x_t|S_1)}{P(x_t|S_2)}\right] = \sum_{t=1}^{T} \ln[P(x_t|S_1)] - \ln[P(x_t|S_2)]$$

Choose S_1 if LLR > 0, S_2 if LLR <0

$$\sum_{t=1}^{T} \ln[P(x_t|S_1)] - \ln[P(x_t|S_2)]$$



Gold & Shadlen (2007) *Ann Rev Neurosci.*



Ludwig & Davies (2011) *Cognitive Psychology.*

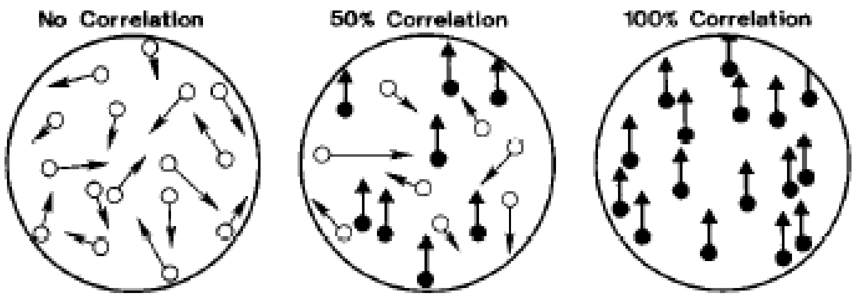
- Sequential sampling is a form of dynamic signal detection theory
- SDT is a useful starting point
- For more details on sequential sampling: Chris on Thursday

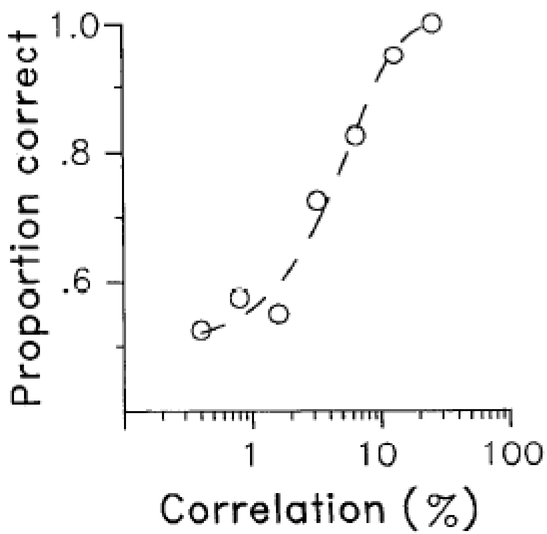
Signal Detection Theory account of choice accuracy

The Journal of Neuroscience, December 1992, *12*(12): 4745–4765

The Analysis of Visual Motion: A Comparison of Neuronal and Psychophysical Performance

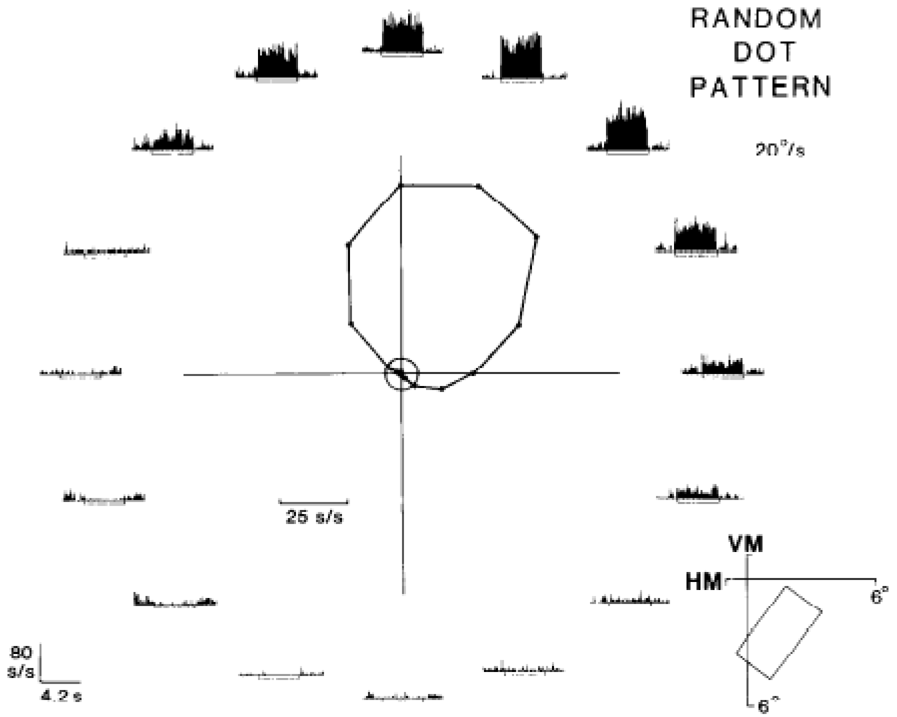
Kenneth H. Britten,[1] Michael N. Shadlen,[1] William T. Newsome,[1] and J. Anthony Movshon[2]





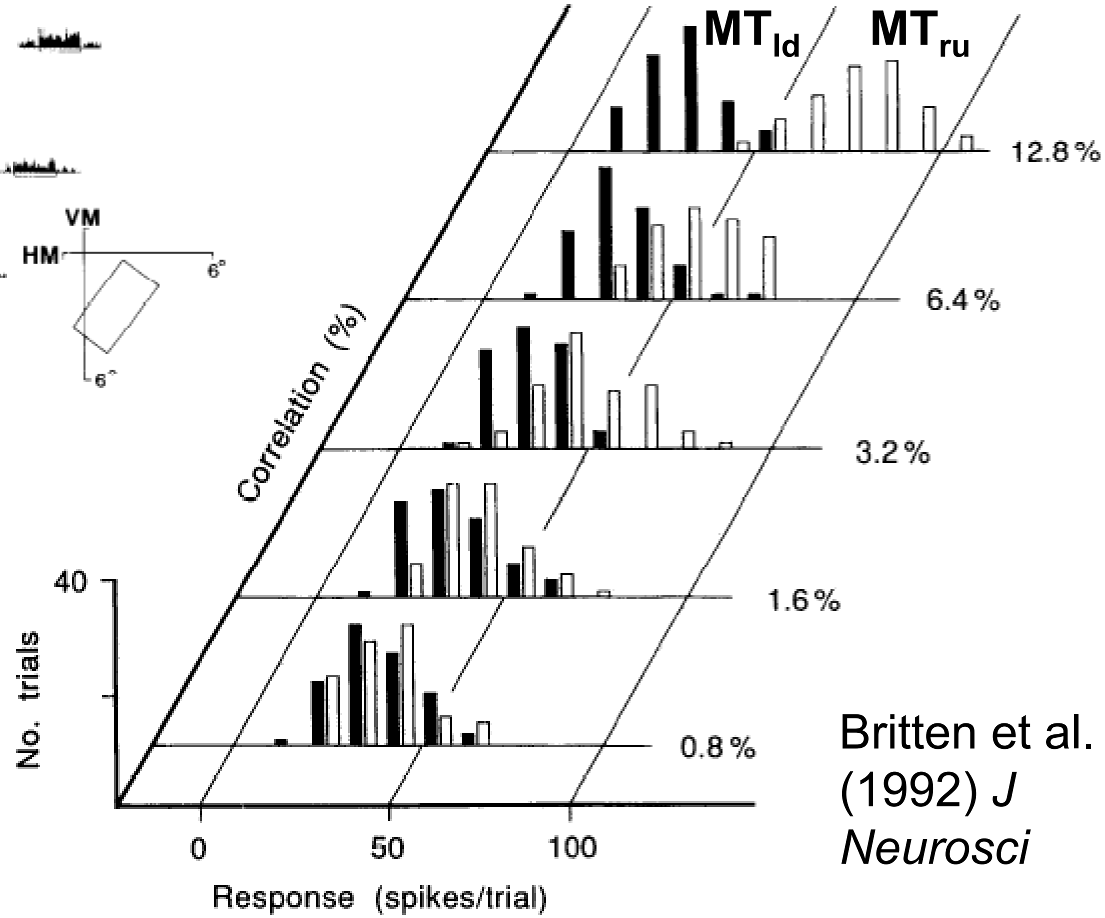
If:

$MT_{ru} > MT_{ld}$: "right and up"
$MT_{ld} > MT_{ru}$: "left and down"

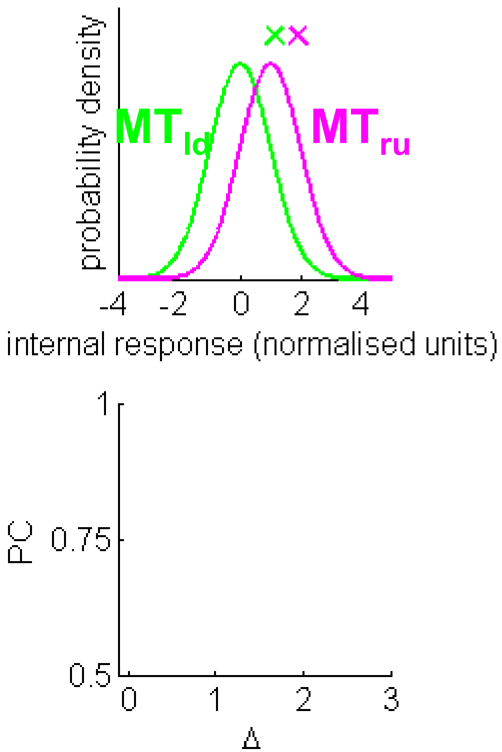


Albright (1984) *J Neurophys*

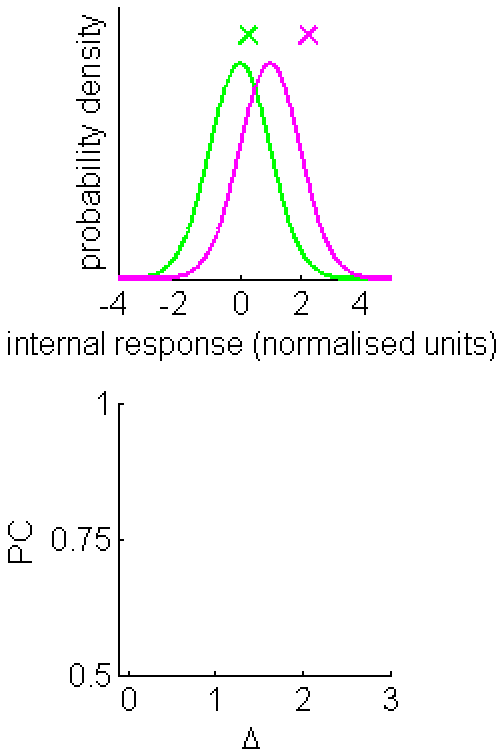
MAX decision rule

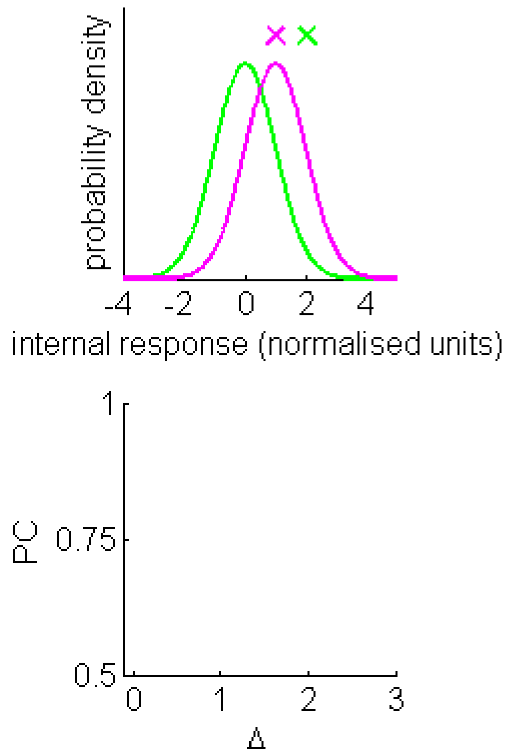


Britten et al. (1992) *J Neurosci*

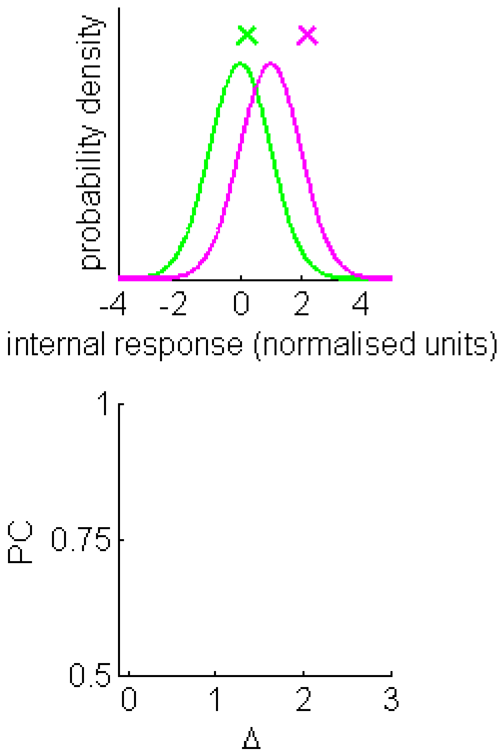


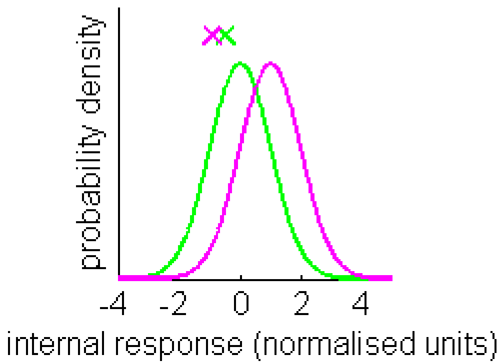
probability density
MT$_{ld}$
MT$_{ru}$
-4 -2 0 2 4
internal response (normalised units)
PC
1
0.75
0.5
0 1 2 3
Δ

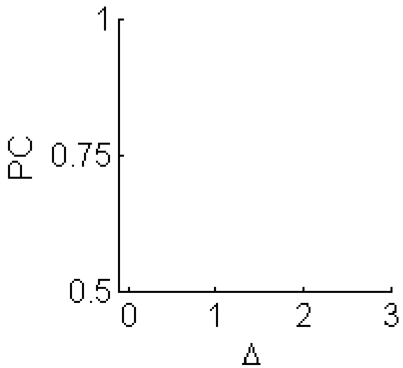

probability density
-4
-2
0
2
4
internal response (normalised units)
PC
1
0.75
0.5
0
1
2
3
Δ

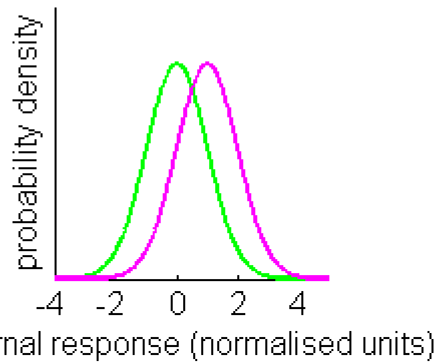


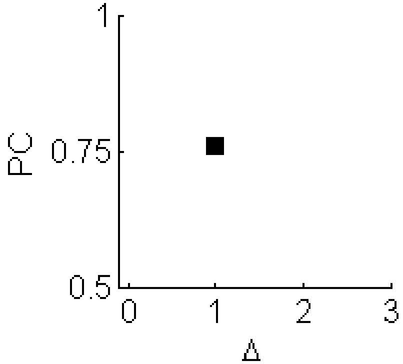
probability density
-4
-2
0
2
4
internal response (normalised units)
1
0.75
0.5
PC
0
1
2
3
Δ


probability density
-4
-2
0
2
4
internal response (normalised units)
1
0.75
0.5
PC
0
1
2
3
Δ


probability density
-4
-2
0
2
4
internal response (normalised units)


1
PC
0.75
0.5
0
1
2
3
Δ


probability density
-4
-2
0
2
4
internal response (normalised units)


1
PC
0.75
0.5
0
1
2
3
Δ

Play time...

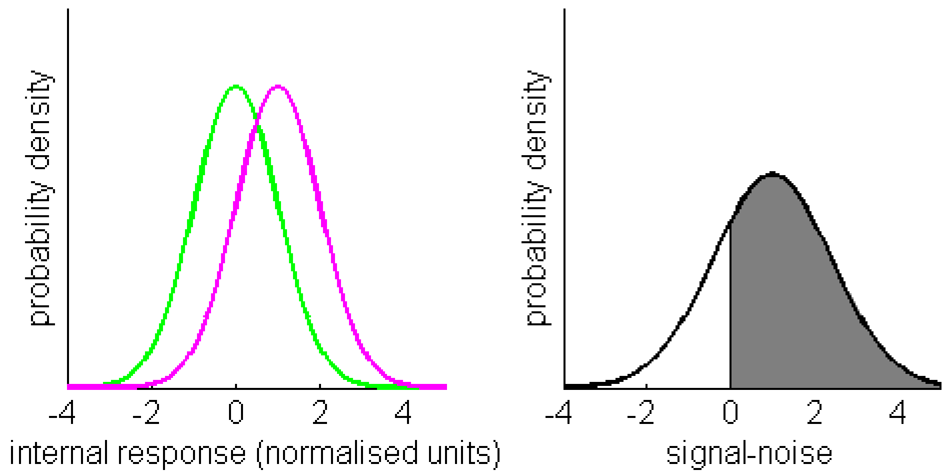
- Simulate pairs of responses: one from the "preferred direction" (signal) distribution, one from the "non-preferred direction" (noise) distribution
- If signal > noise, response is correct
- Fix noise arbitrarily (e.g. set to 1)
- Vary the separation between the two distributions at a number of values (e.g. motion coherence): e.g. 0, 0.5, 1, etc.
- Plot PC as a function of separation

From internal responses to accuracy

- Method 1: simulation
  - Draw N pairs of responses, one response from signal and noise distributions (N = large)
  - Evaluate # trials on which $X_{signal} > X_{noise}$
  - For ties, flip a coin – I use: rbinom
  - PC = #/N
  - Generally not needed for parametric models (i.e. where we assume some known distributional form)
  - For example, where we have ***measured*** two distributions of noisy responses (like in Britten et al., 1992)

From internal responses to accuracy

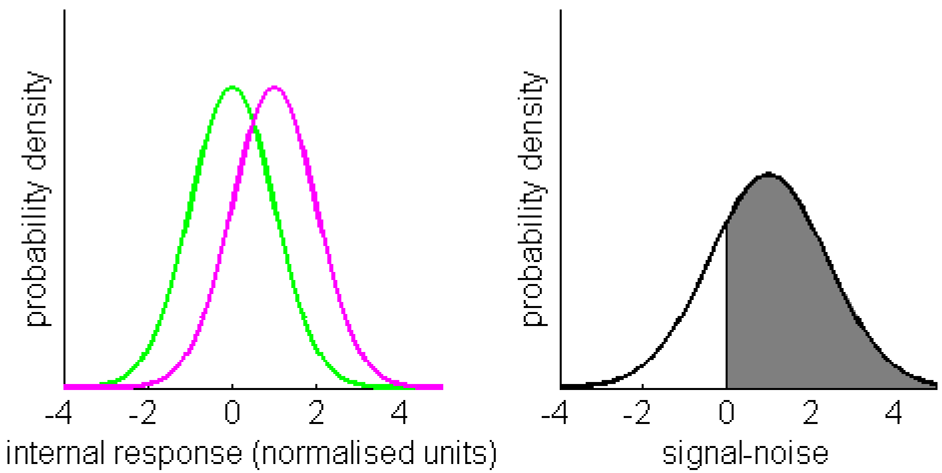
- Method 2: difference-of-Gaussians
  - o MAX rule corresponds to evaluating whether (X_{signal} – X_{noise}) > 0



$$\mu_{difference} = \mu_{signal} - \mu_{noise} \qquad \sigma_{difference} = \sqrt{\sigma^2_{signal} + \sigma^2_{noise}}$$

From internal responses to accuracy

- Method 2: difference-of-Gaussians
  - Compute area under curve > 0
  - = 1 – area up to 0
  - In R this is easy: 1-pnorm(0, mean=$\mu_{\text{difference}}$, sd=$\sigma_{\text{difference}}$)

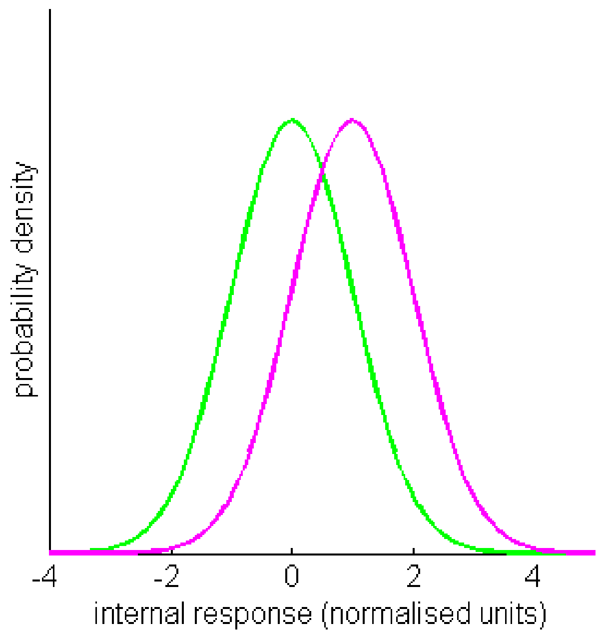


From internal responses to accuracy

- Method 3: direct evaluation
  - For a given noise value $u = X_{noise}$, compute $P(X_{signal} > u)$
  - Weigh $P(X_{signal} > u)$ by the probability of observing u under the noise distribution
  - Do this for all possible values for u and add up, i.e. ***integrate*** across u
  - Think of this as the “average probability that the signal response was greater than the noise response”
  - For 2-AFC this is equivalent to difference-of-Gaussian method
  - Direct evaluation is necessary for M-AFC (M > 2)

From internal responses to accuracy

- Method 3: direct evaluation
  - o Signal and noise probability density: $f_s(x)$ $f_n(x)$



$$P(X_{signal} > u) = \int_u^{\infty} f_s(x)dx = H(u)$$

$$PC = \int_{-\infty}^{\infty} f_n(u)H(u)du$$

From internal responses to accuracy

- Method 3: direct evaluation

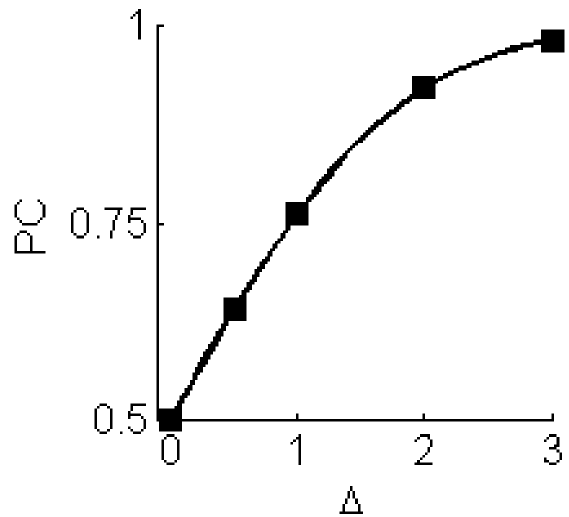
$$PC = \int_{-\infty}^{\infty} f_n(u)H(u)du$$

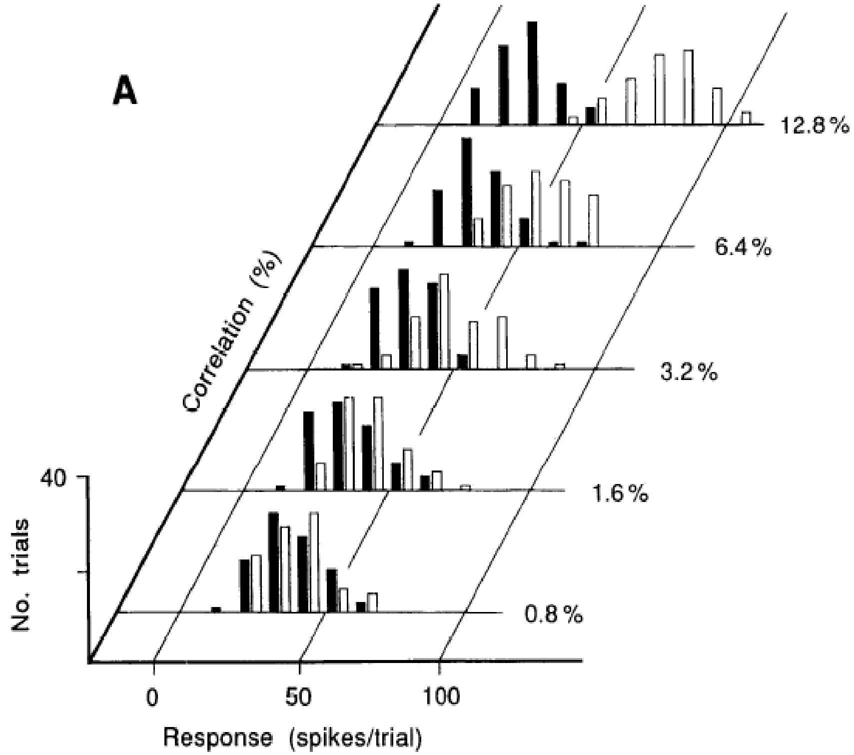
  - o PC cannot be directly evaluated
  - o Gaussian distributions have to be integrated numerically (i.e. approximated with a discrete sum)
  - o In R:

```
# Method 3: Integrate across all possible noise responses and evaluate the
# probability that the signal response was greater. This results in a more
# complex expression to integrate, but the result should be identical to
# the difference distribution. The advantage of this approach is that it
# scales up for designs with more than two alternatives.
# First, set up the function to integrate
TwoAFCmax=function(x,mean1=0,mean2=0,sd1=1,sd2=1){
  y=dnorm(x,mean=mean1,sd=sd1)*(1-pnorm(x,mean=mean2,sd=sd2))
  return(y)
}
# This function can only be numerically integrated. You can be more or less
# sophisticated about this. I'm using the built-in 'integrate'
# function. Other functions to explore would be 'quadgk' or, the simplest
# of all, 'trapz'.
lowlim=mu_noise-5*sigma_noise # Set some sensible integration limits
uplim=mu_signal+5*sigma_signal # Assumes: mu_noise <= mu_signal
PC=integrate(TwoAFCmax,
             lowlim,
             uplim,
             mean1=mu_noise,
             mean2=mu_signal,
             sd1=sigma_noise,
             sd2=sigma_signal)
PCint=PC$value
```

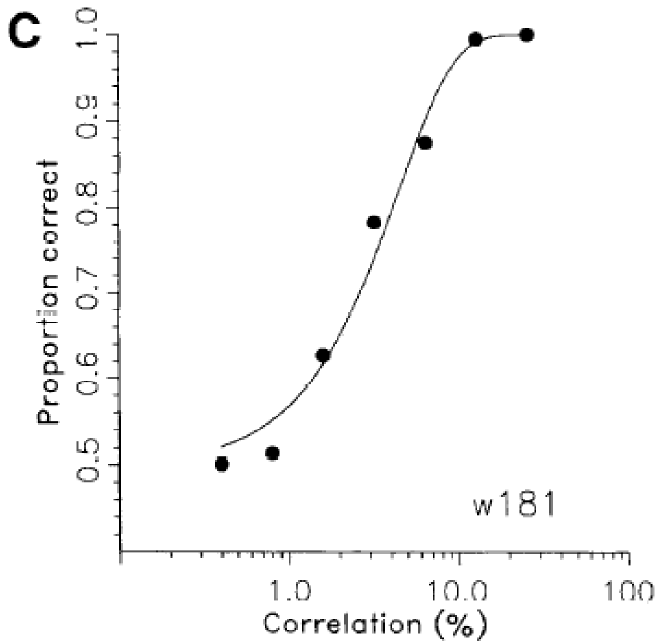
R script:
TwoAFC_accuracy_demo

(Run with mu_signal = e.g. 0, 0.5, 1, 2, 3—plot in one window)

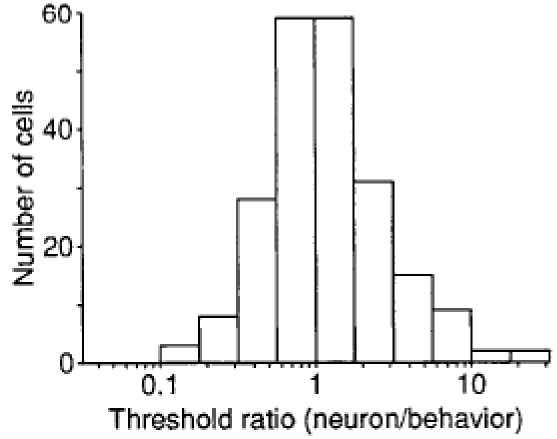


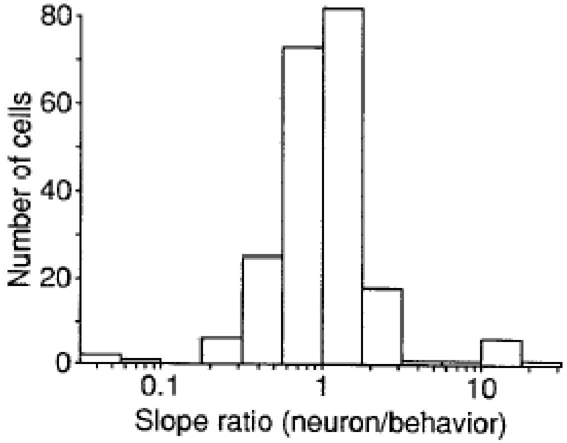


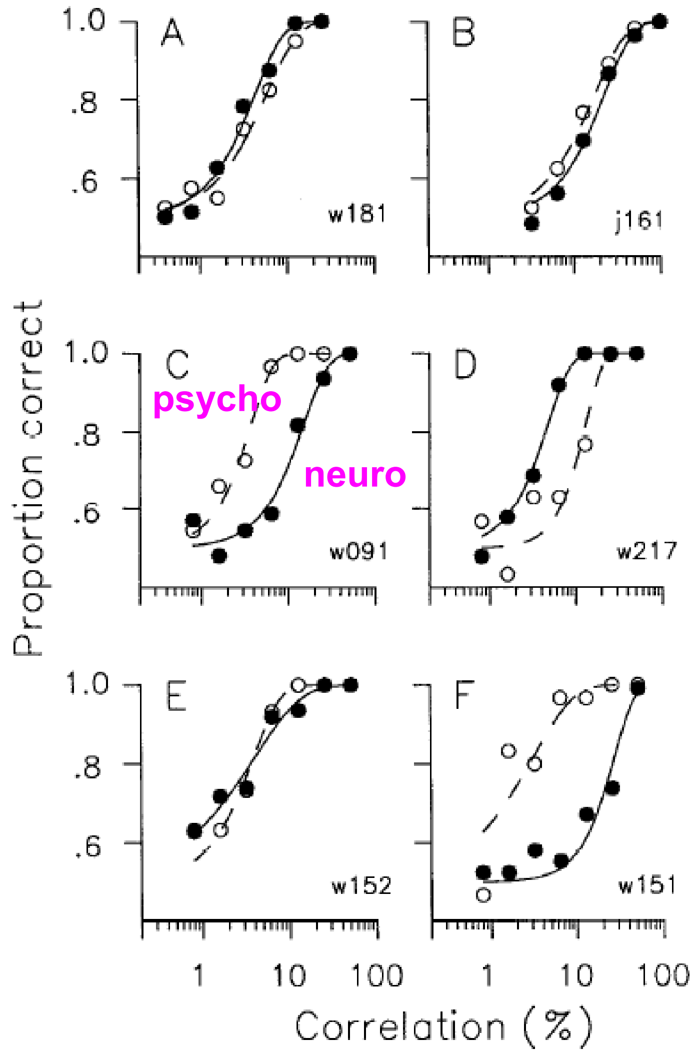
Britten et al. (1992) *J Neurosci*



Psychometric – neurometric comparison







Aim

- Develop a mechanistic model of the psychometric function
- Fit model to behavioural data using MLE
- or use Bayes if you like :-)
- Fit data on a training set
- Outcome metric: predictive generalisation to test set (i.e. likelihood of the “un-seen” test data)
- It may be instructive to compute other metrics for model selection (e.g. AIC, BIC)

The data set

Cognitive Psychology 63 (2011) 61–92

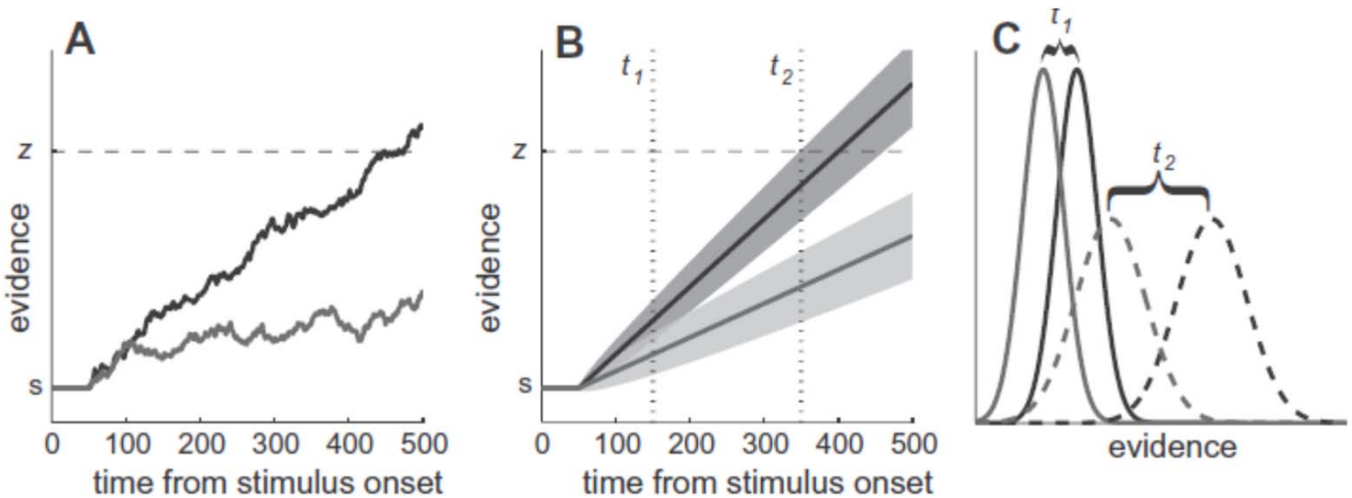
Contents lists available at ScienceDirect

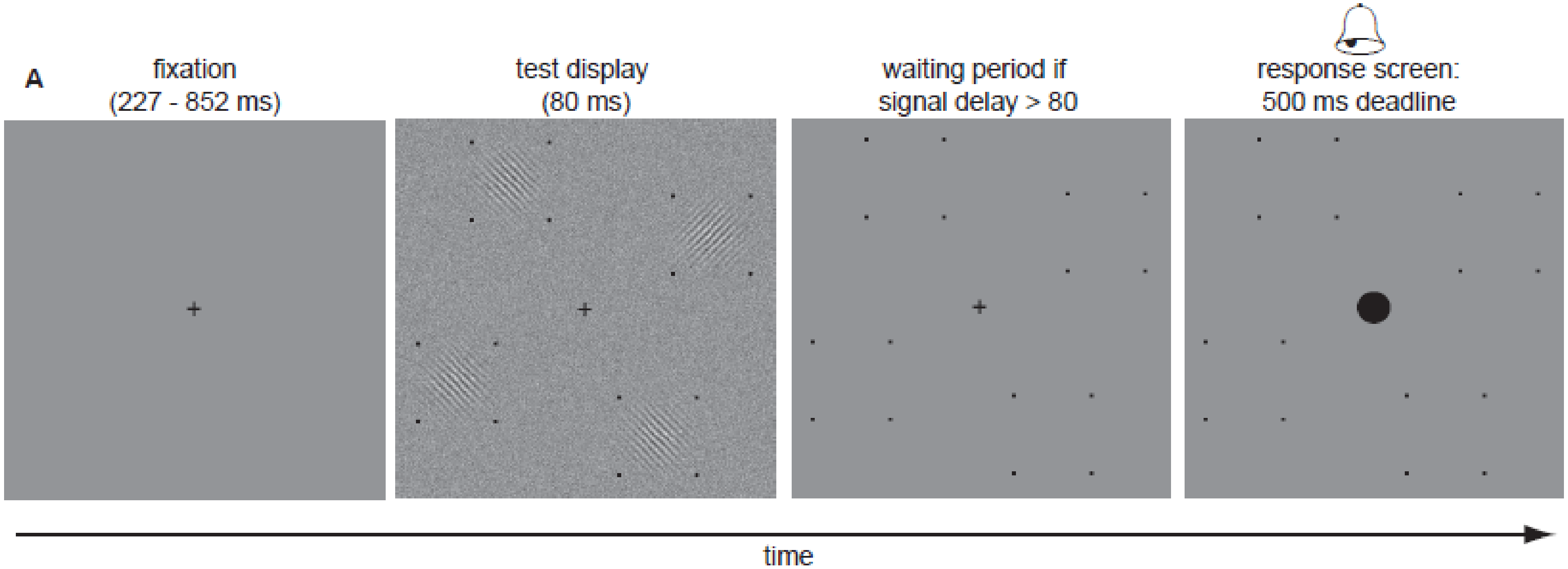
Cognitive Psychology

journal homepage: www.elsevier.com/locate/cogpsych

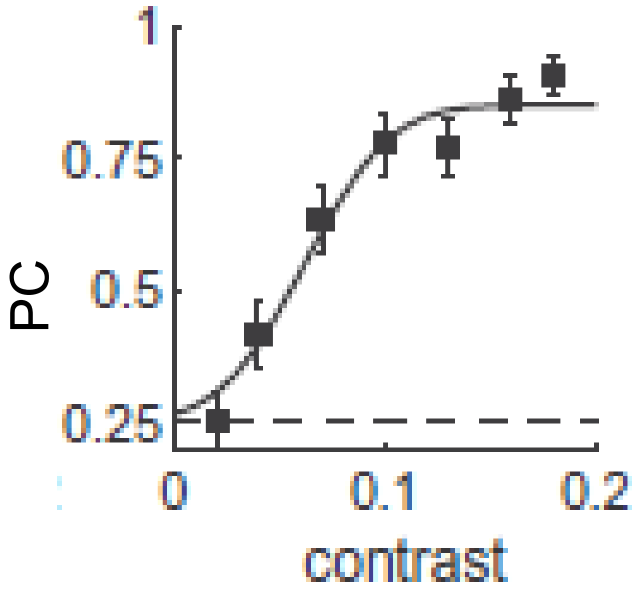
Estimating the growth of internal evidence guiding perceptual decisions

Casimir J.H. Ludwig *, J. Rhys Davies





- Search for oriented target among 3 distractors (orthogonal orientation)
- Manual localisation response
- Vary luminance contrast of all four patterns (energy)



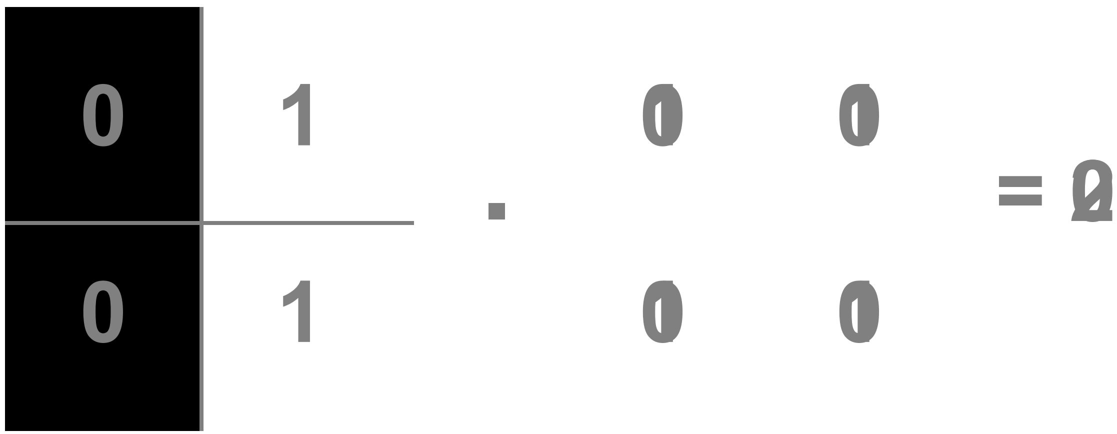
- PC(c) = *F*(c; **θ**), where **θ** is a vector of model parameters
- In this case: *F* = cumulative Gaussian, **θ** = {μ, σ}
- *Function* fitting

Aim: fit a "process" model

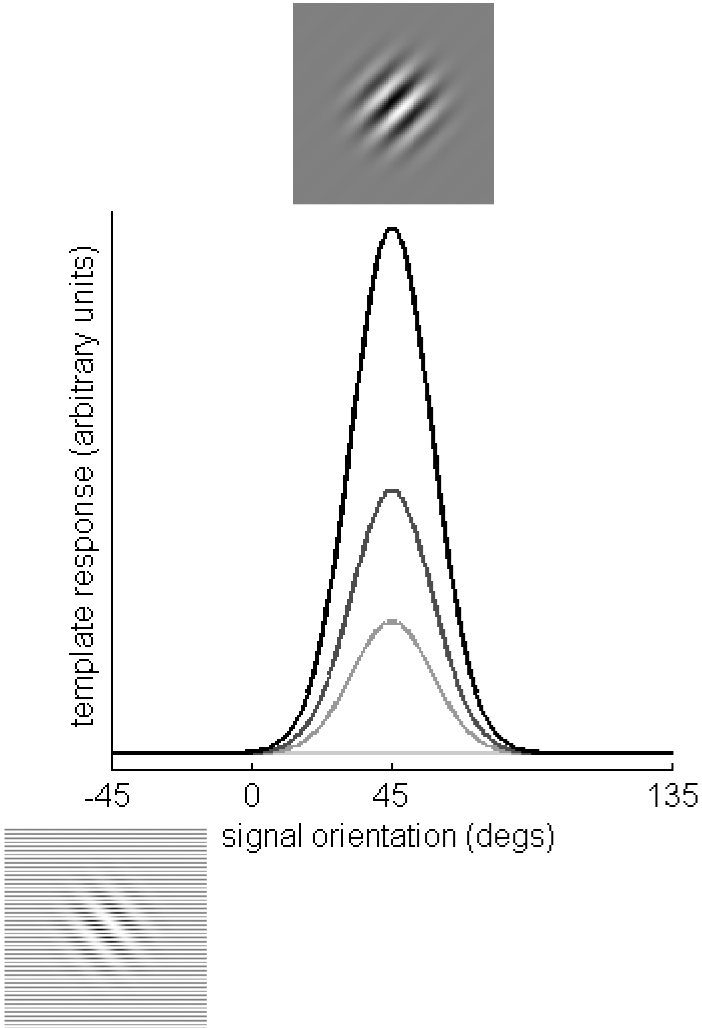
- Template matching!
- Implemented as 2-D "correlation" between a "target template" *T* and pattern *S* on screen

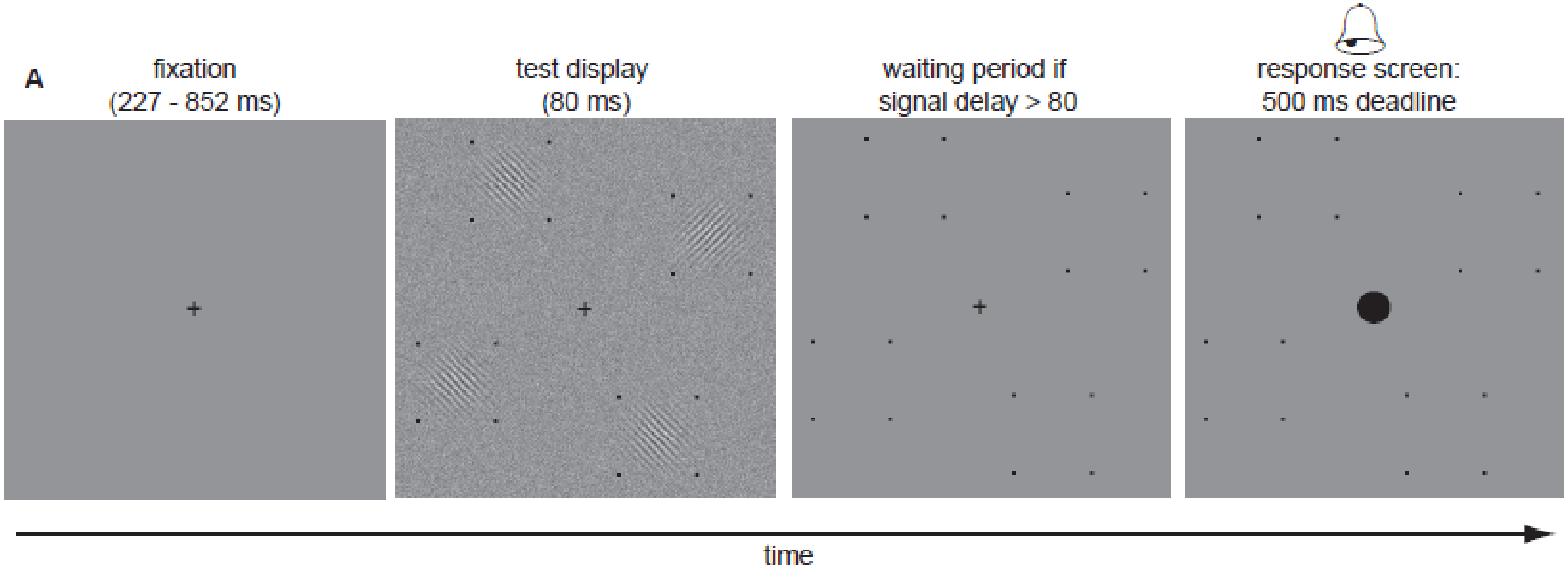
$$r = \sum_{x,y} T(x,y) \times S(x,y) = T(x,y) \cdot S(x,y)$$

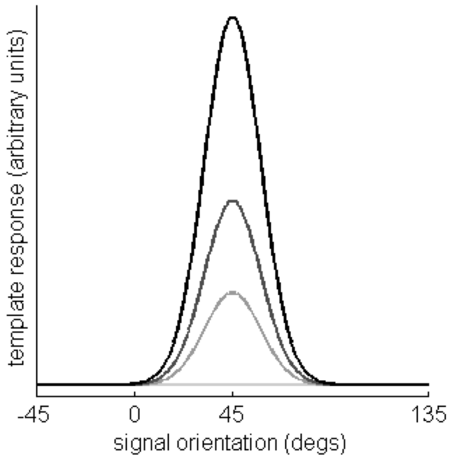
- Images are just numbers!

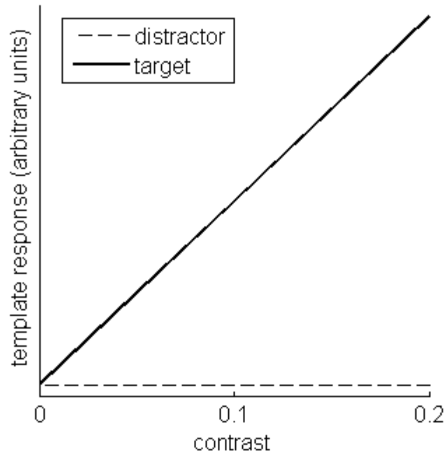


Best template matches the signal exactly ("matched filter")

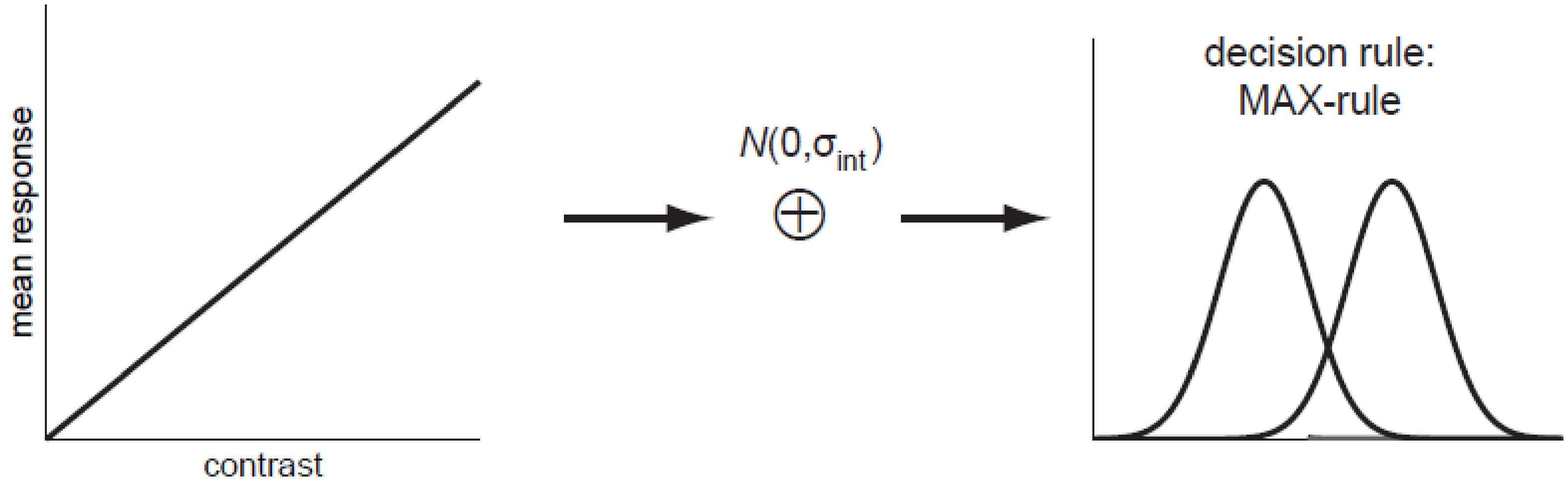



A
fixation
(227 - 852 ms)
test display
(80 ms)
waiting period if
signal delay > 80
response screen:
500 ms deadline
time


template response (arbitrary units)
-45
0
45
135
signal orientation (degs)


template response (arbitrary units)
distractor
target
0
0.1
0.2
contrast

A minimal model



- Template matched at all 4 locations
- Distractor response = 0
- Target response > 0
- Internal noise corrupts internal responses
- MAX response = most likely target location

- Similar to model used for random dot motion
- Except:
- 4 alternatives, rather than 2
- Continuous link function between signal strength (contrast here) and mean internal response

M-AFC

- Difference-of-Gaussian method cannot be applied when M > 2
- Direct evaluation:
  - o For a given target response $u = X_{signal}$, compute probability that ***all*** (M-1) distractor responses are less than u:

$$P(X_{n1} < u) \times P(X_{n2} < u) \times \ldots \times P(X_{nM-1} < u) = \left[\int_{-\infty}^{u} f_n(x)dx \right]^{M-1}$$

- Weigh P(all X_{noise} < u) by the likelihood of observing u under the target response distribution
- Do this for all possible values for u and add up, i.e. ***integrate*** across u

$$\boxed{\mathrm{dnorm}(u,\mathrm{mean}=\mu_{signal},\mathrm{sd}=\sigma_{signal})}$$

$$PC = \int_{-\infty}^{\infty} f_s(u) \left[\int_{-\infty}^{u} f_n(x)dx \right]^{M-1} du$$

- Think of this as the "average probability that the target response was greater than all three distractor responses"

$$\boxed{\mathrm{pnorm}(u,\mathrm{mean}=\mu_{noise},\mathrm{sd}=\sigma_{noise})}$$

$$P(X_{n1} < u) \times P(X_{n2} < u) \times \ldots \times P(X_{nM-1} < u) = \left[\int_{-\infty}^{u} f_n(x)dx \right]^{M-1}$$

Model in a nutshell

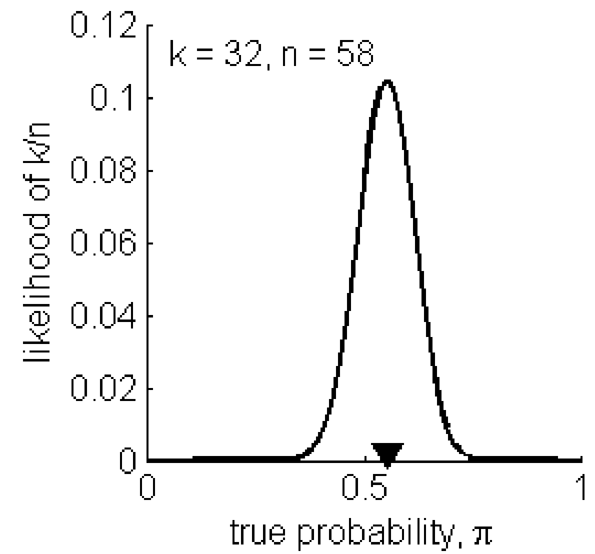
- Matched filtering:
  - Linear transducer $\mu_{\text{target}}(c) = \beta c$
  - Zero response to distractors $\mu_{\text{nontarget}}(c) = 0$
- We cannot separate out signal/noise, so fix noise arbitrarily: $\sigma_{\text{nontarget}} = \sigma_{\text{target}} = 1$
- Free parameters: β

Fitting the model

- In R, load ‘TrainData.csv’
- Dataframe with 5 variables: participant, contrast, square root signal energy, number correct responses, total number of trials
- Number of contrast levels and trials may vary across participants

Fitting the model

- For each signal level: *n* bernoulli trials (i.e. binary outcome), with *k* out of *n* successes
- E.g. *k* = 32, *n* = 58
- Binomial likelihood function:



Fitting the model

- Observed data are *frequencies* and the model predicts the “true” probability, π, from which those frequencies were generated

$$\mathrm{L}(\pi;n,k)=\frac{n!}{k!(n-k)!}\pi^{k}(1-\pi)^{n-k}$$

- Or, in R: $\mathrm{dbinom}(k,n,\pi)$
- Look at your SIMPLE code from last week

Fitting the model

- Fit model to each participant separately
  - For each signal: $\beta E \longrightarrow \pi \longrightarrow$ likelihood
  - Across signals multiply likelihoods or, rather, ***sum log-likelihoods***
  - Multiply by -2 (Deviance)
  - Minimise (e.g. optim or whatever optimisation routine you want) to find best fitting parameter(s)
- Overall goodness-of-fit: sum Deviance measures across participants

Improving the model

- Non-linear transducer?
  - o Logarithmic: $\mu(c) = \beta \ln c$
  - o Power function: $\mu(c) = (\beta c)^{\gamma}$
  - o Naka-Rushton equation (often used for single neuron response functions—ask me)
- Signal-dependent noise: $\sigma^2 = 1 + k\mu$
- Mixture of visually guided responses and stimulus-independent responses (“finger errors”, λ): $PC_i = w\pi_i + (1-w)\lambda$

Model selection

- Winner is determined with a common yard stick:
  - o Fix the parameters of your best model fit to the training data
  - o Compute total deviance for the test set
- If you have (lots of) time to spare:
  - o Play around with other model selection methods: e.g. AIC, BIC, n-fold cross-validation (on the training and test sets combined)
  - o Do different fitting and model selection methods give you the same answer?

Tournament!

Final considerations & comments

- What is the number of data points (for BIC)?
- What model selection method is best?
- Is MAX rule optimal?
- Applications beyond psychophysics?

Final considerations & comments

- What is the number of data points (for BIC)?
  - Total number of trials for a subject (i.e. 100s in this case per subject)?
  - Number of "proportion" samples (typically <10 per subject)?
- Similar considerations in more "typical" psychology experiments
  - E.g. N subjects doing M trials in K conditions
  - Model average RT in K conditions
  - N x K?
  - N x M x K?

Final considerations & comments

- What model selection method is best?
- With a bit of luck, all selection methods say the same thing!
- Goals of the modelling:
  - Best account of one or more benchmark data sets?
  - Generating predictions for future behaviour or experimental effects?
- Cross-validation makes no assumptions about measuring/quantifying complexity
- Very natural assessment of over-fitting/generalisation, but no meaningful scale

Final considerations & comments

- Is MAX rule optimal?
  - Given four noisy pattern responses X_i for i=1,…,4
  - At each location *i*: is the response more likely to have been generated from the “target distribution”?
  - Compute LR – Location with greatest LR is most likely to have contained the target

$$LR_i = \frac{f(X_i \mid \mu_{\text{target}}, \sigma_{\text{target}})}{f(X_i \mid \mu_{\text{nontarget}}, \sigma_{\text{nontarget}})}$$

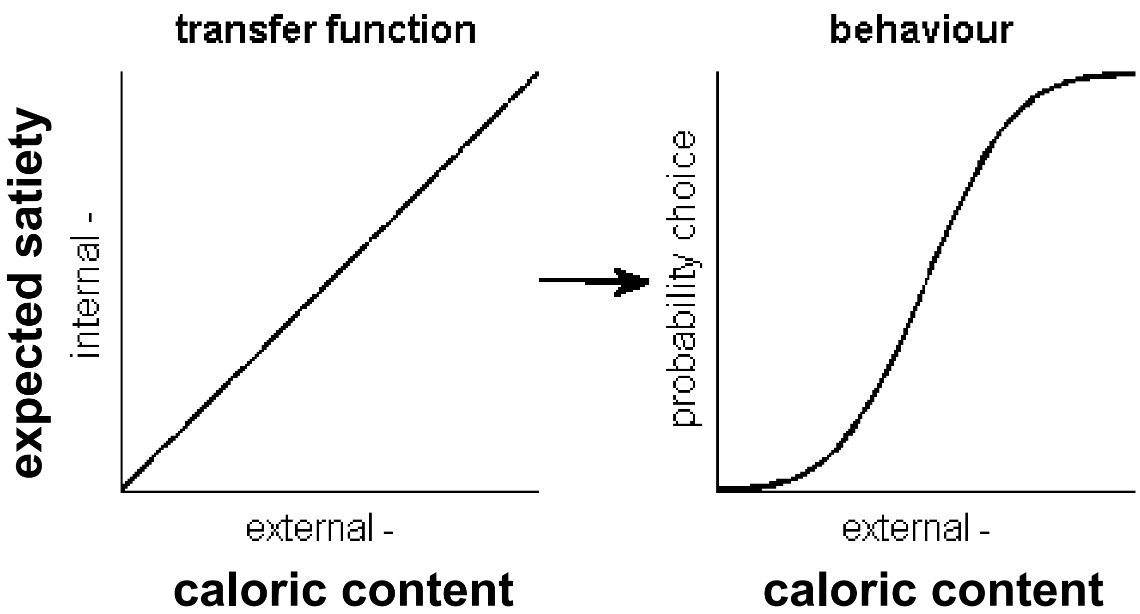
(Recall that $\mu_{\text{target}} \geq \mu_{\text{nontarget}} = 0$, $\sigma_{\text{target}} = \sigma_{\text{nontarget}} = 1$, and assume equal prior probability across locations)

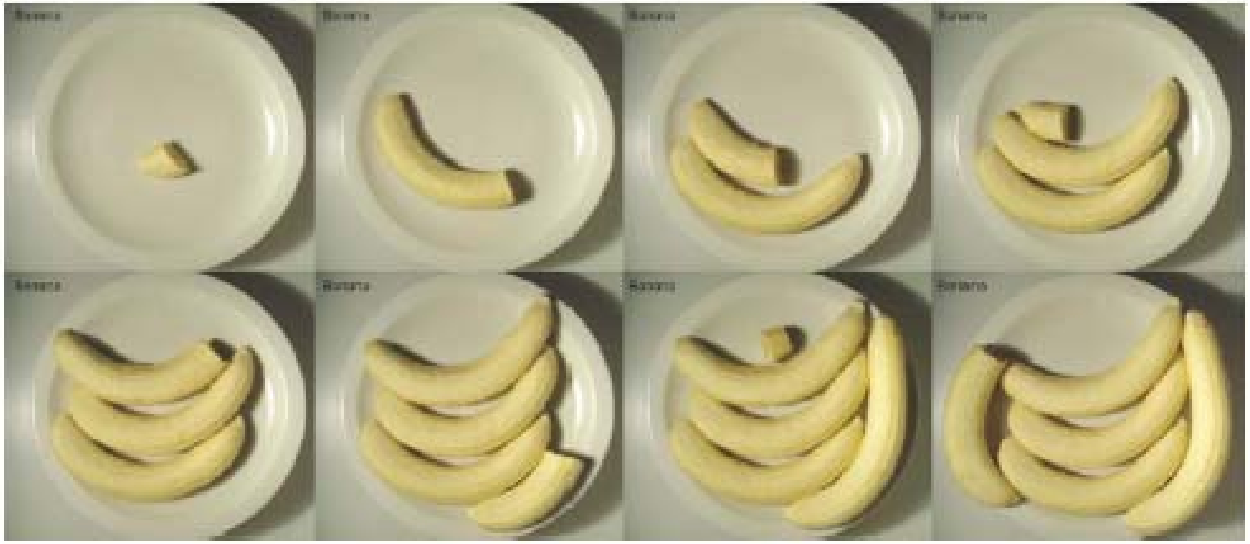
Final considerations & comments

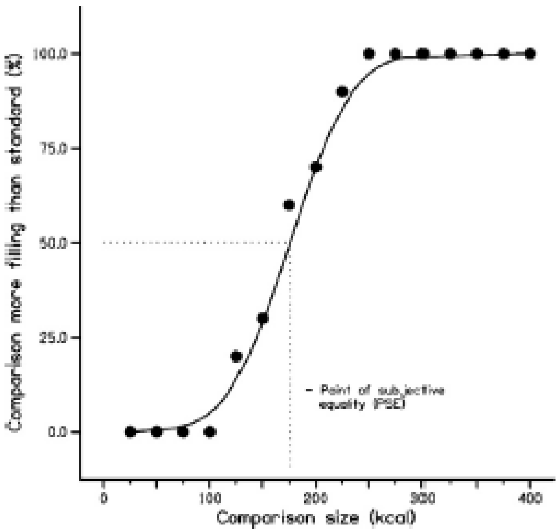
- Applications beyond psychophysics?
- Three key ingredients:
  - o Transfer function relating external dimension to an internal variable
  - o Noise
  - o Choice rule

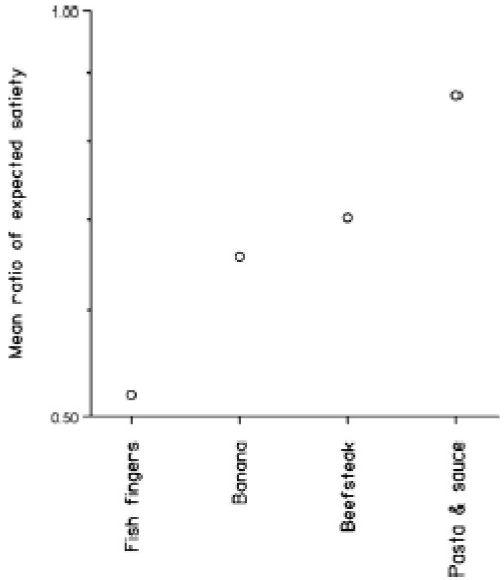
Final considerations & comments

- Applications beyond psychophysics?





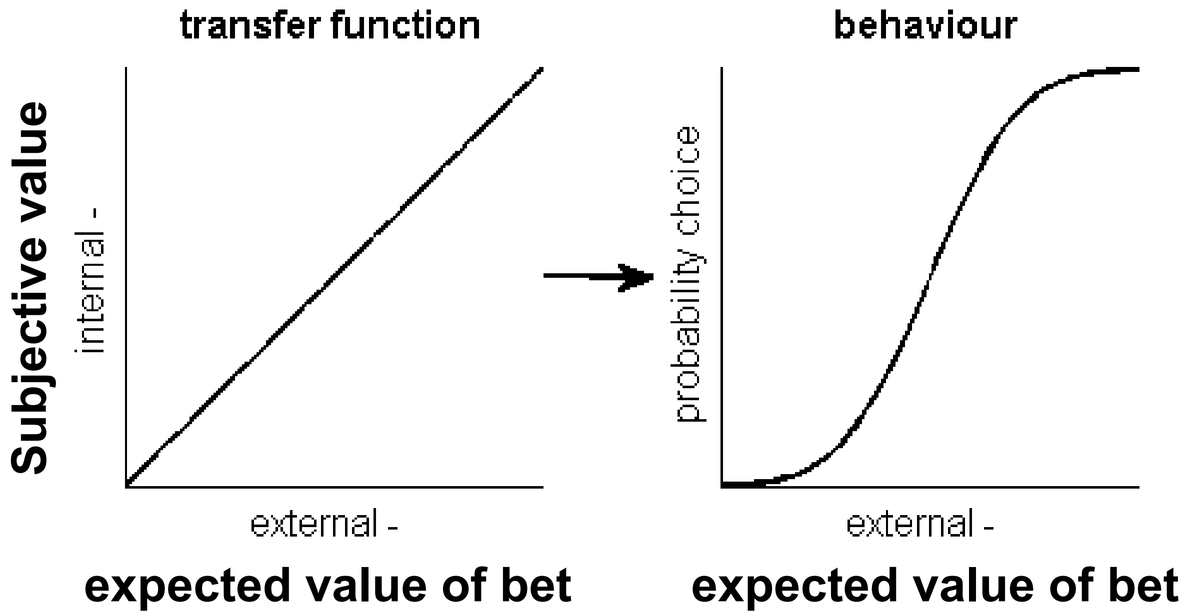




Brunstrom et al. (2008) *Appetite*

Final considerations & comments

- Applications beyond psychophysics?



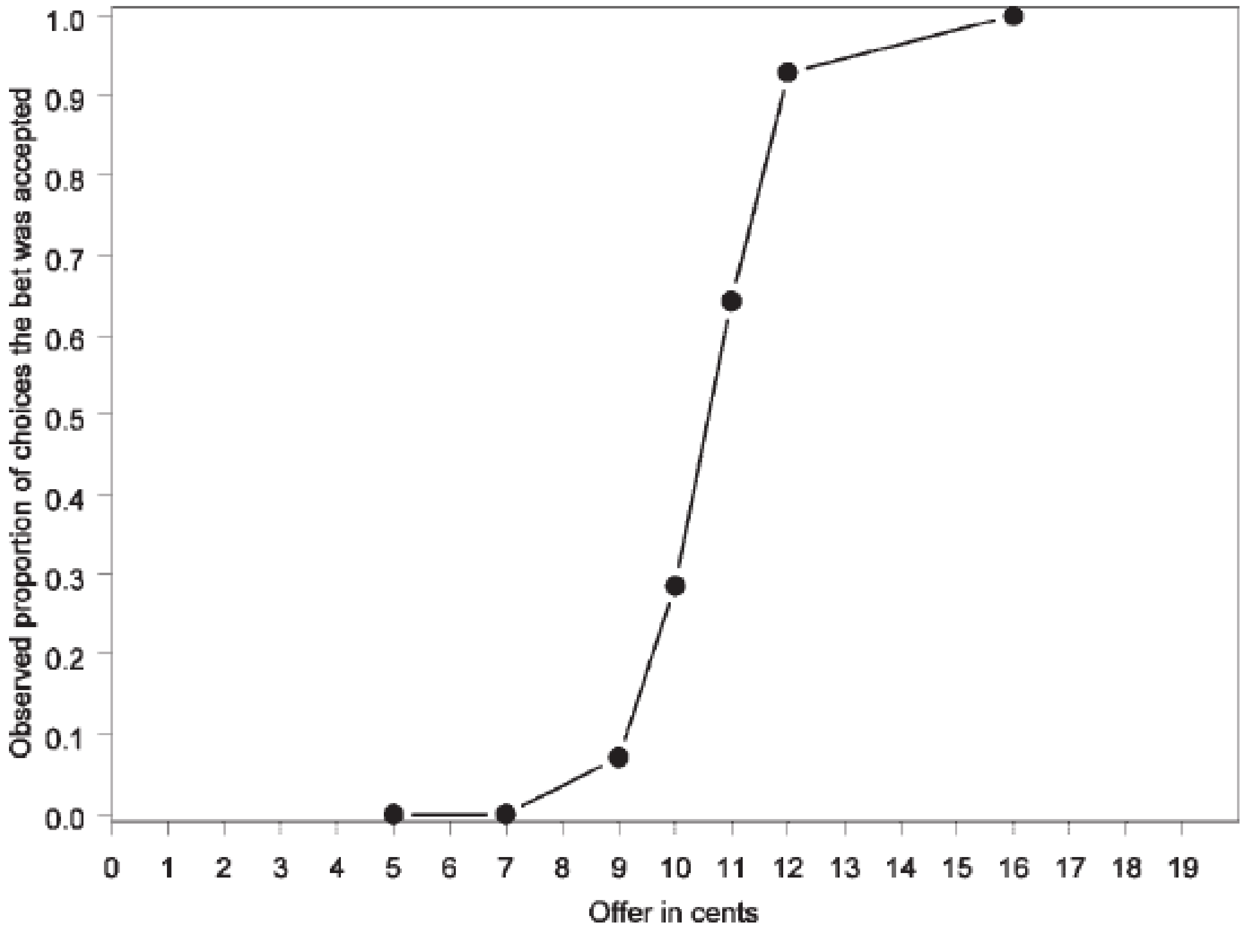
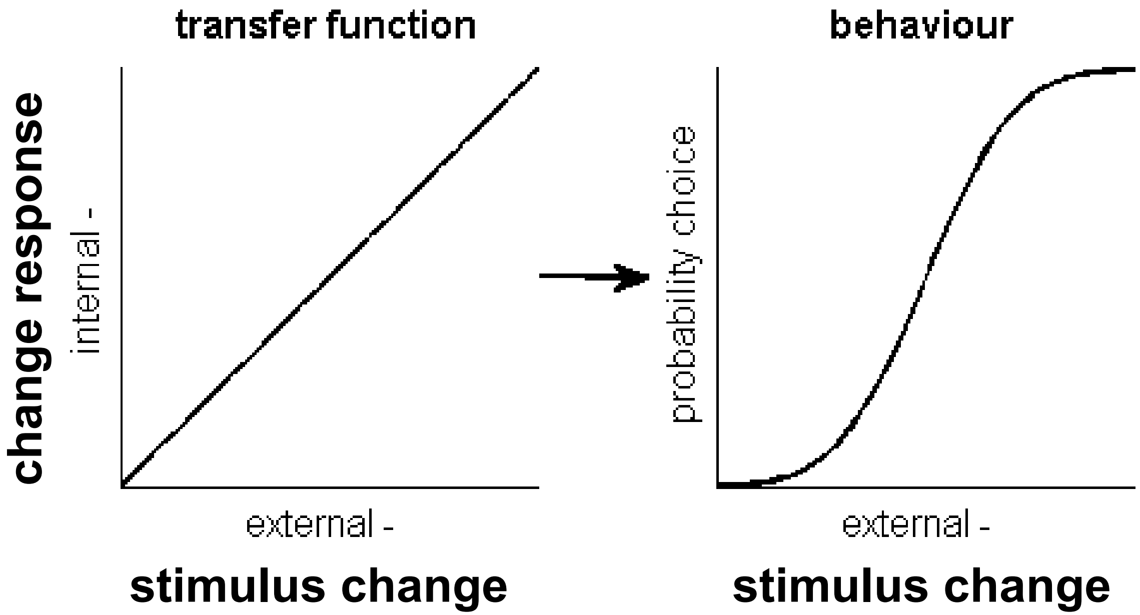


Figure 1. The choice proportions of participant B-I in Mosteller and Nogee (1951) of accepting a bet that led with a probability of .33 to a specific gain represented on the abscissa and with a probability of .67 to a loss of 5 cents. Each bet was offered 14 times, and when the participant rejected the bet, a payoff of zero resulted.

Mosteller & Nogee (1951), from Rieskamp (2008) *JEP: LMC*

Final considerations & comments

- Applications beyond psychophysics?



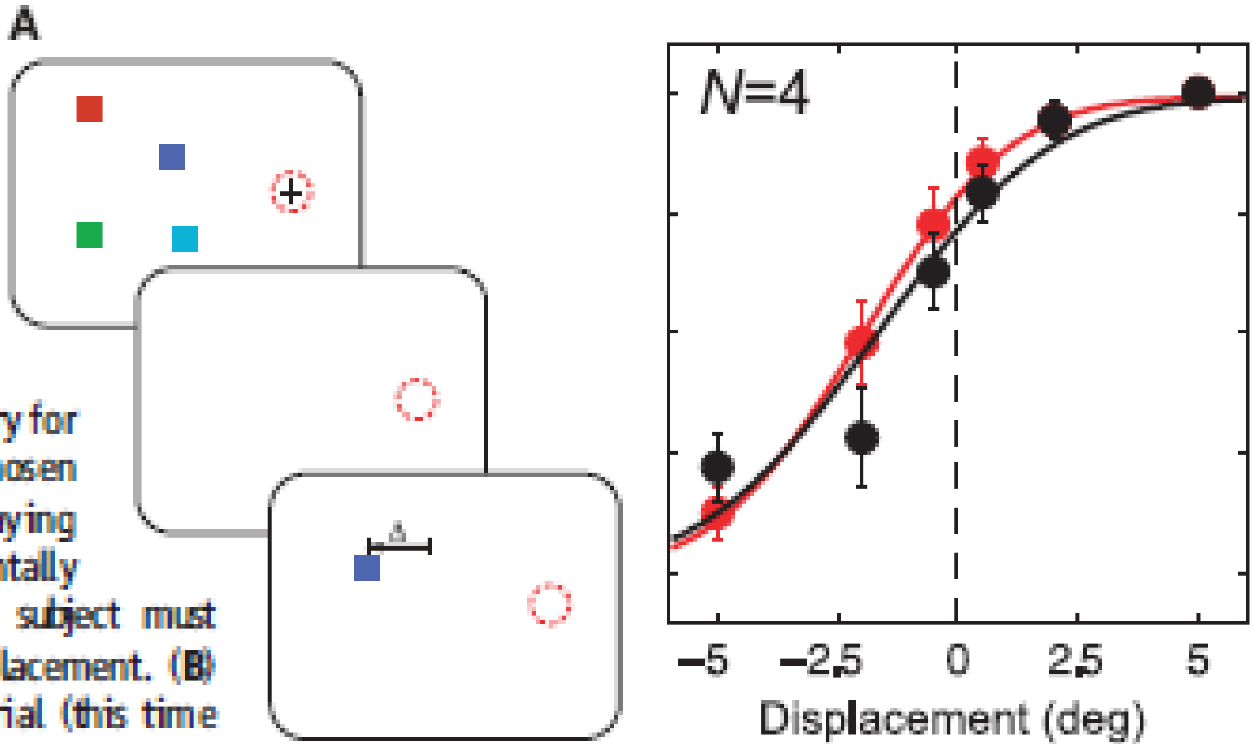


Fig. 1. Experimental procedure. **(A)** Stimuli and sequence of events on a location-judgment trial. The example shown is a fixation trial with a set size of four items. After the sample display is blanked, subjects' memory for location of a randomly chosen item is tested by redisplaying the item displaced horizontally through distance Δ. The subject must report the direction of displacement. **(B)** An orientation judgment trial (this time shown for the saccade condition, with a set size of two items). At the tone, the subject makes a saccade toward the item of a prespecified color (here green) with the display being blanked during the eye movement. A randomly chosen item is redisplayed, rotated through an angle Δ, and the subject reports the direction of rotation. Red circles indicate gaze position.

Bays & Husein (2008) *Science*