ECVP2025

Technische Universität Berlin

Workshop

MLDS and MLCM: two scaling methods to study stimulus appearance

Guillermo Aguilar

Technische Universität Berlin

ECVP 2025 Mainz

24th August 2025

Website

Detailed Conference Program

WLAN: UNI-MAINZ

Identity: **ecvp25**
PW: **#JohannesGutenberg**
System certificates
Domain: **uni-mainz.de**

ecvp2025.uni-mainz.de

ecvp2025@uni-mainz.de

Contents – 13:30 to 16:00

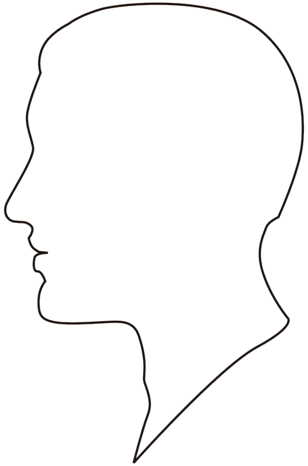
- General introduction to scaling methods
- **Maximum Likelihood Difference Scaling (MLDS)**
  - Theory
  - Exercise: acquire your own data
  - Code: analyze data with the MLDS package
  - Exercise: analyze your data
- **Maximum Likelihood Conjoint Measurement (MLCM)**
  - Theory
  - Code: analyze data with the MLCM package
  - Exercise: acquire and analyze your own data

Intro: psychophysical methods

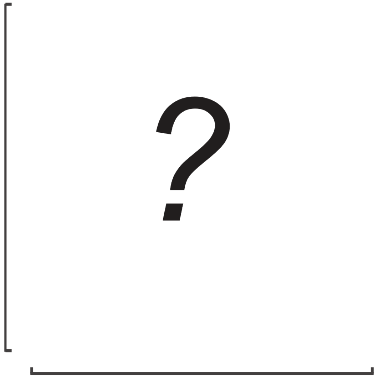
External World

Subjective experience

x → $\Psi(x)$



$\Psi(x)$



x

Intro: psychophysical methods

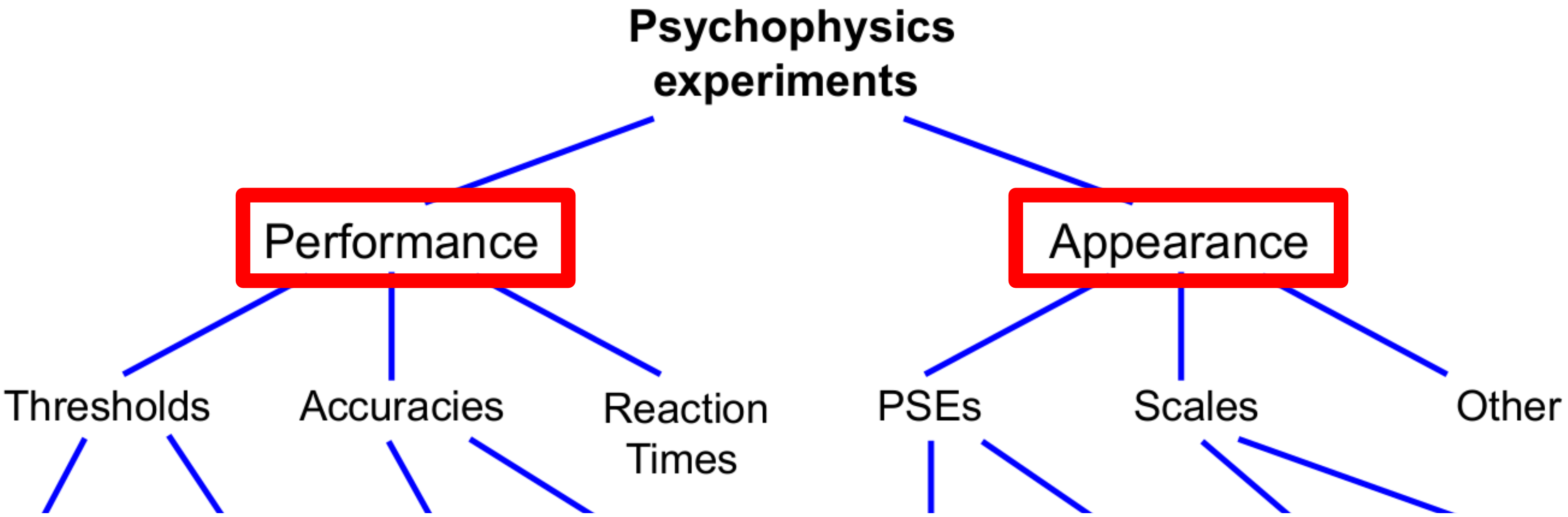
Example: Glossiness (Cheeseman et al. 2021)

Minimum ⟵ Specular Reflectance (ρ_s) ⟶ Maximum

How does specular reflectance map to perceived glossiness?

Physical attribute → Perceptual dimension

Intro: psychophysical methods



e.g.

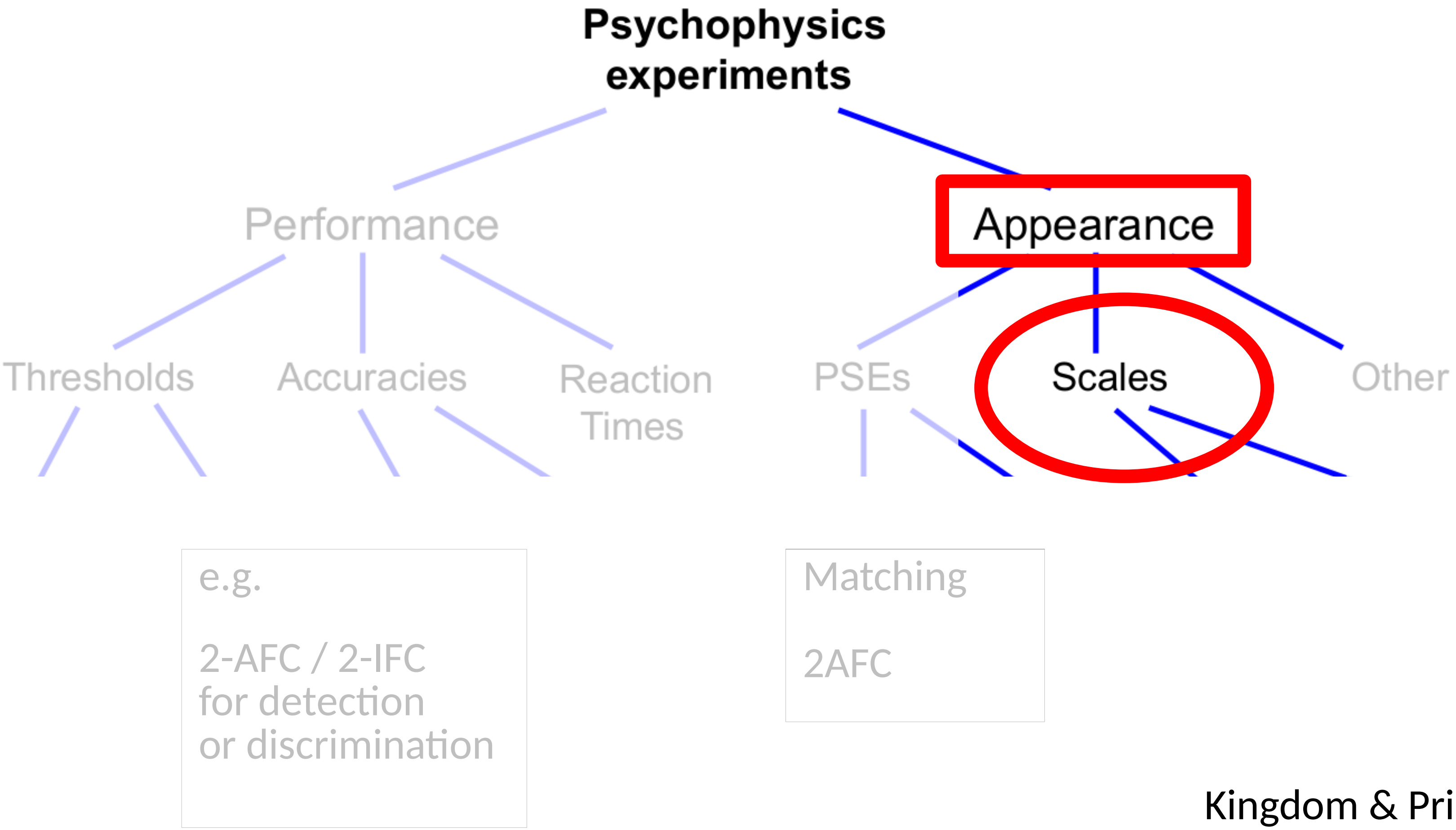
2-AFC / 2-IFC
for detection
or discrimination

Matching

2AFC

Kingdom & Prins (2016)

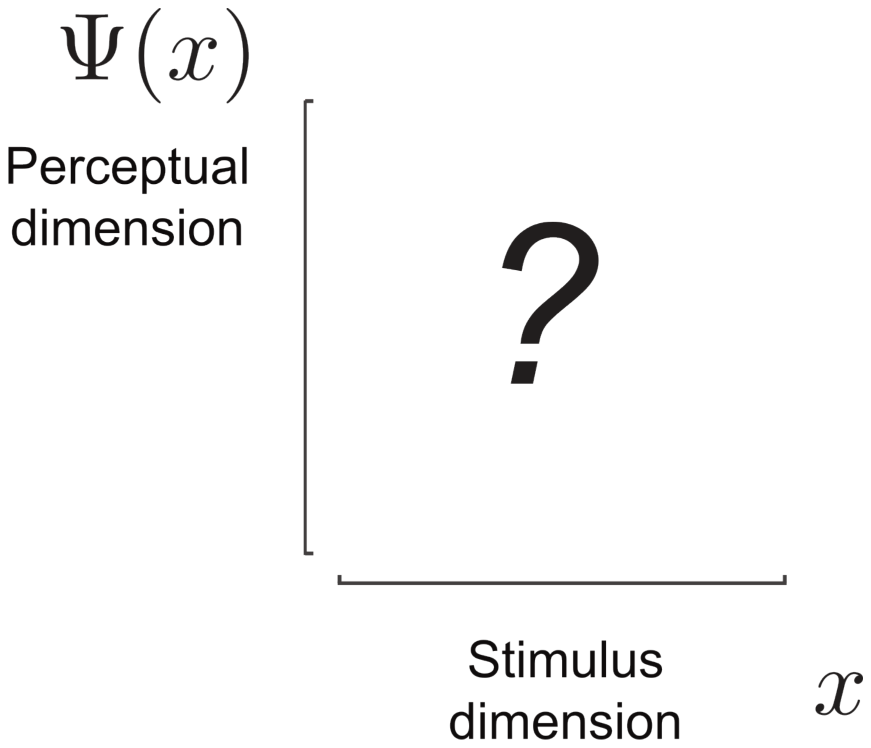
Intro: psychophysical methods



Kingdom & Prins (2016)

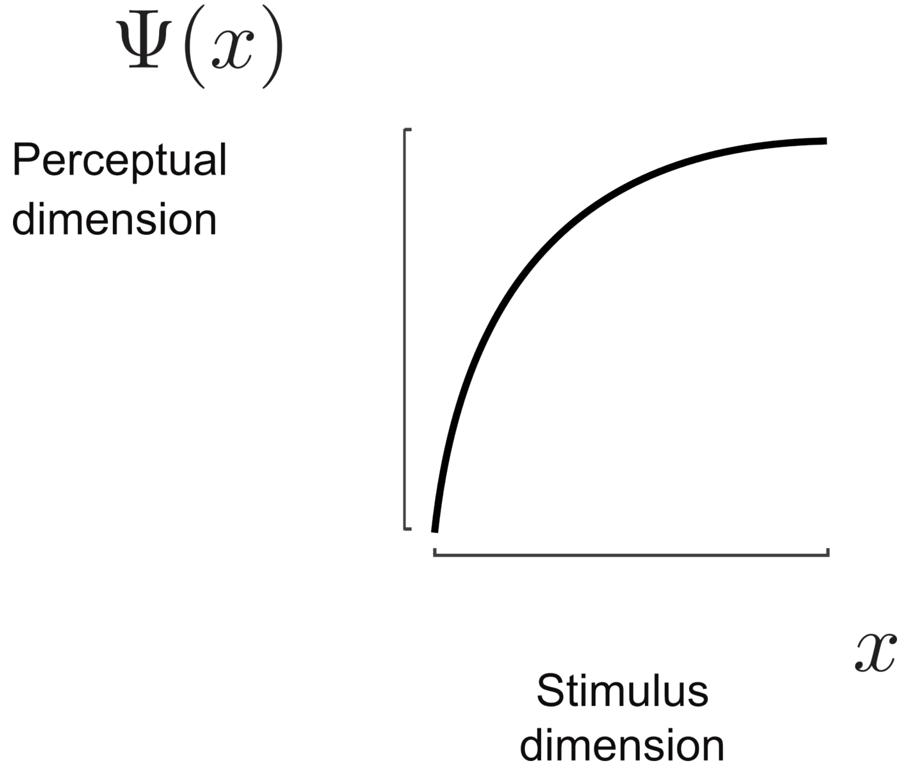
Intro: scaling methods

Aim to **characterize mapping between stimulus and perceptual dimensions**



Intro: scaling methods

Aim to **characterize mapping between stimulus and perceptual dimensions**

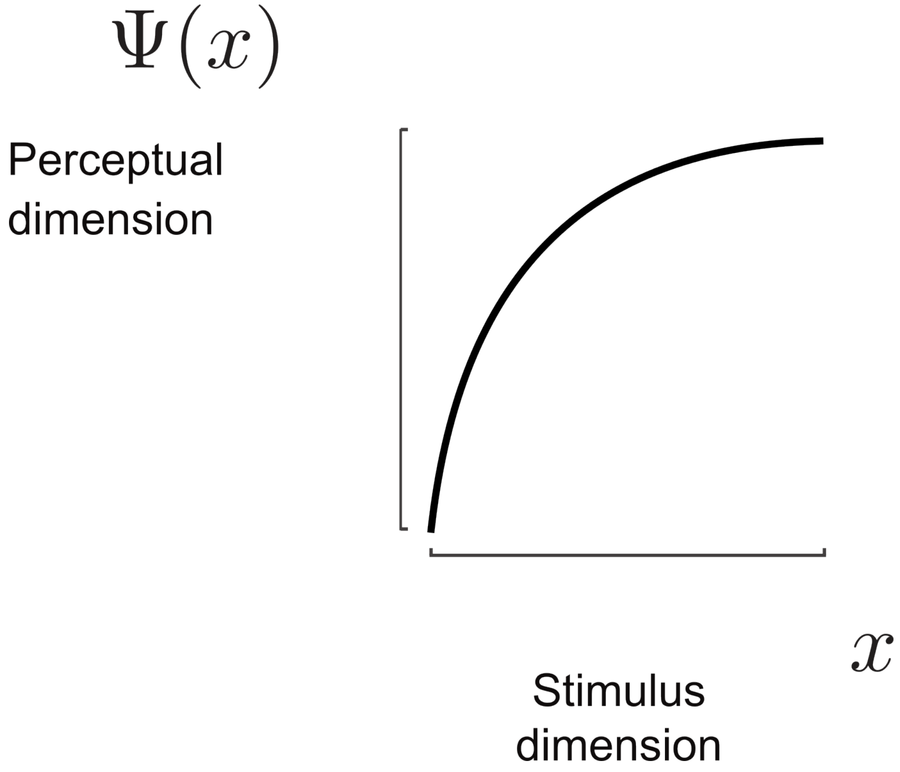


Intro: scaling methods

Aim to **characterize mapping between stimulus and perceptual dimensions**

→ scale (y-axis) should represent the magnitude of perceived intensity

→ make predictions
→ compare with computational models

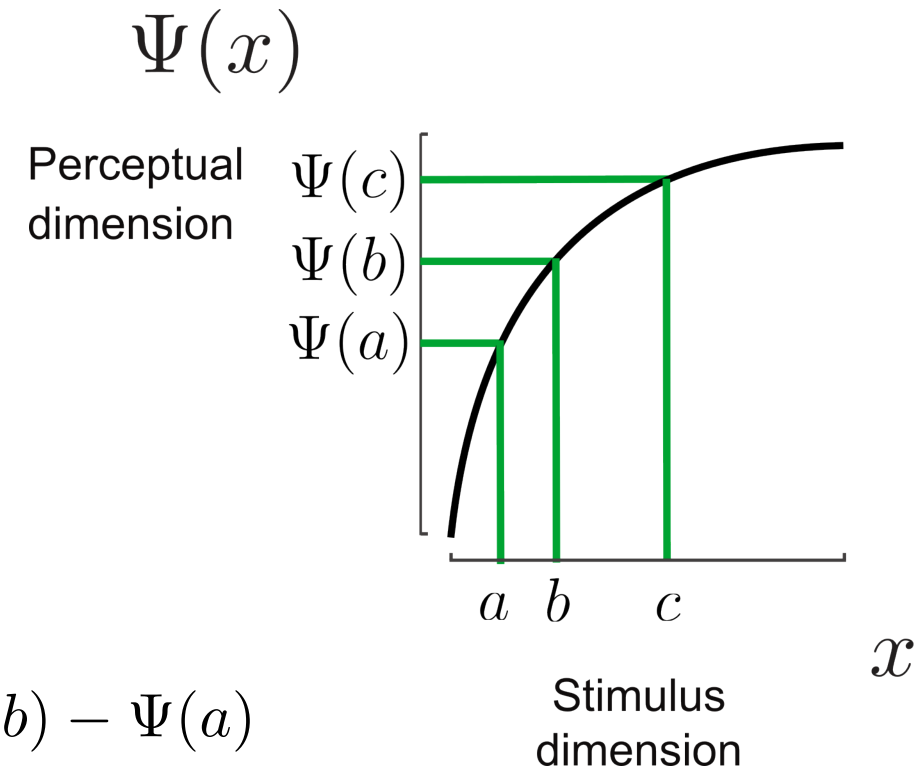


Intro: scaling methods

Aim to **characterize mapping between stimulus and perceptual dimensions**

→ scale (y-axis) should represent the magnitude of perceived intensity

→ make predictions
→ compare with computational models



$$\Psi(c) - \Psi(b) \sim \Psi(b) - \Psi(a)$$
$$(c - b) >> (b - a)$$

First thought: magnitude estimation

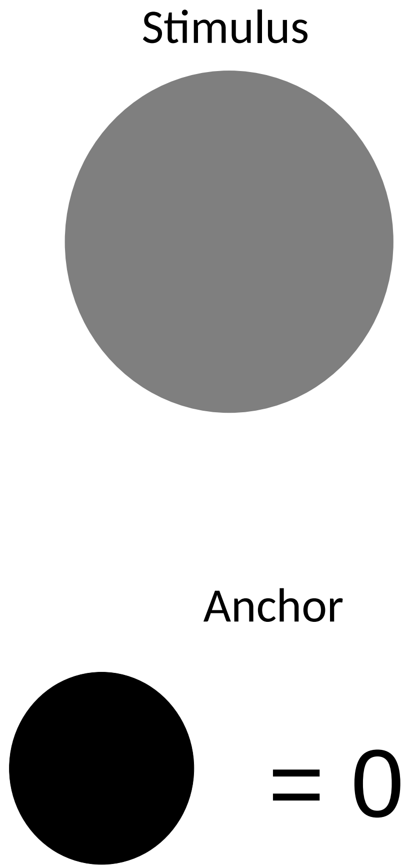
- Task:
  "Give a number that reflects the perceived intensity of the stimulus"

- Known to produce unreliable and suffer from confounds

  - Regression effect
    → O's do not use whole range

  - Anchoring effects
    → scale' shape changes depending on the anchor given

- Assumes linear mapping between perception and the "number line"



Maximum Likelihood Difference Scaling (MLDS)

Key features

- Scaling method based on judgments of **perceptual differences** (intervals)

- For variations on **one perceptual dimension** (e.g. contrast, brightness, glossiness, image quality)

- **Forced choice** task, producing binary data

- Explicit **observer model**

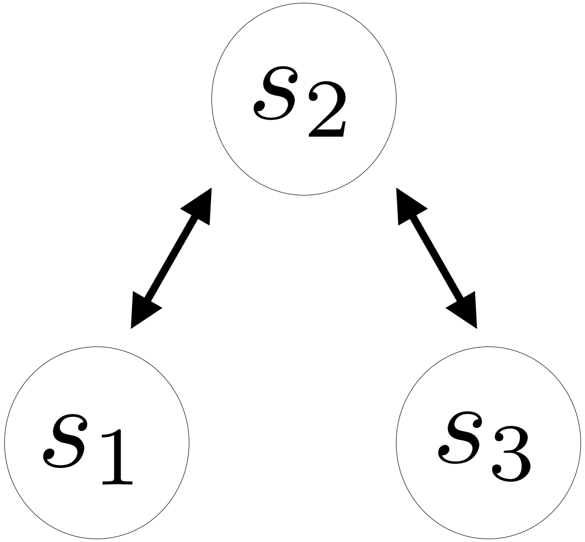
- Model considers **stochasticity** on the observers' judgments

Maloney & Yang (2003)
Knoblauch & Maloney (2012)

Experimental design in MLDS

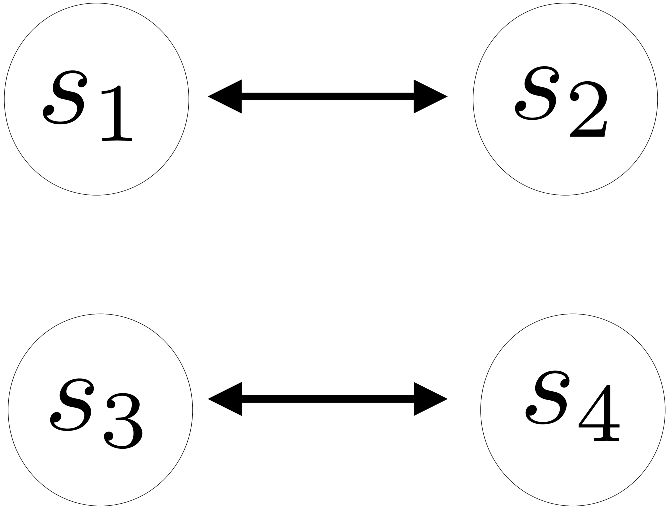
Task: **method of triads or quadruples.**

Triads



Which stimulus is more different
that the middle one?

Quadruples



Which pair is more different?

Forced-choice judgment

Advantages:

- judgment is a binary decision
  → more easily modeled
  → need of an observer model (more detail later)

- quick trials, only 1 – 2 seconds
  (vs. several seconds of adjustment in matching tasks)

Disadvantages:

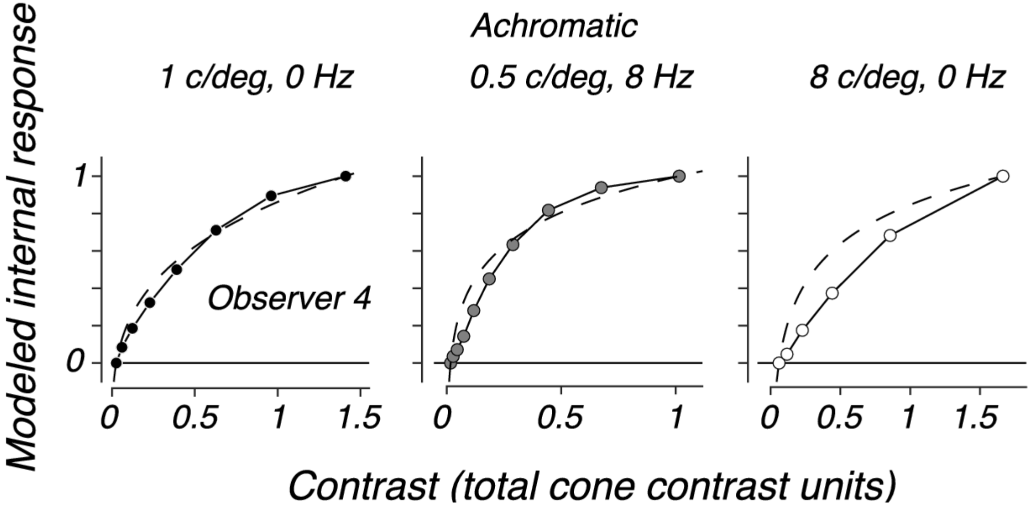
- requires many repetitions

Examples of MLDS usage

Contrast perception (Shooner & Mullen, 2022)

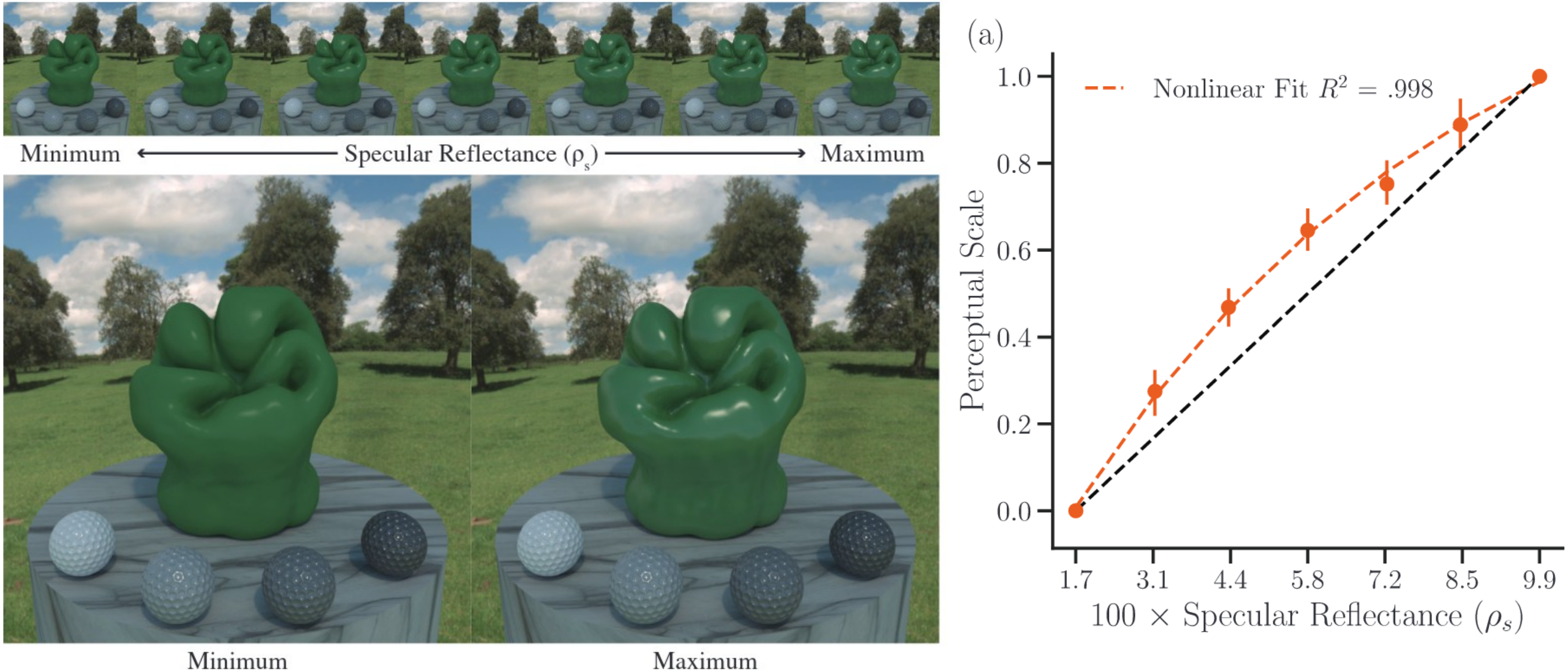
Experiment 2: Contrast difference scaling

"Which of the bottom stimuli differs most from the stimulus on top?"



Examples of MLDS usage

Glossiness (Cheeseman et al. 2021)

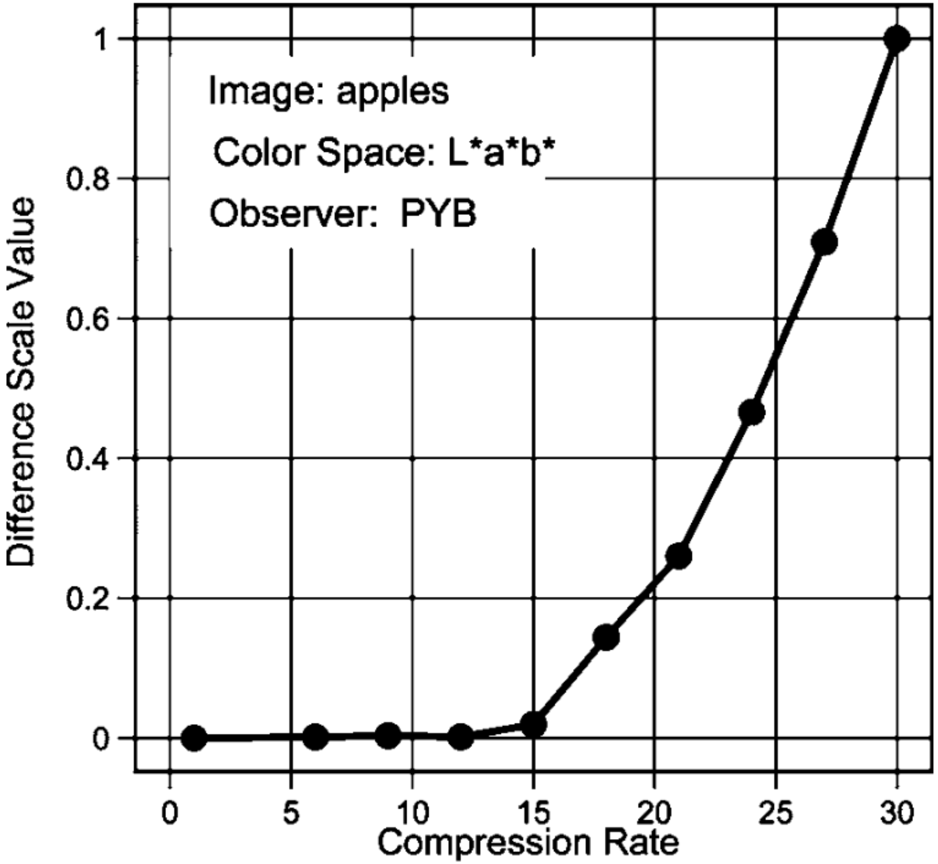


Examples of MLDS usage

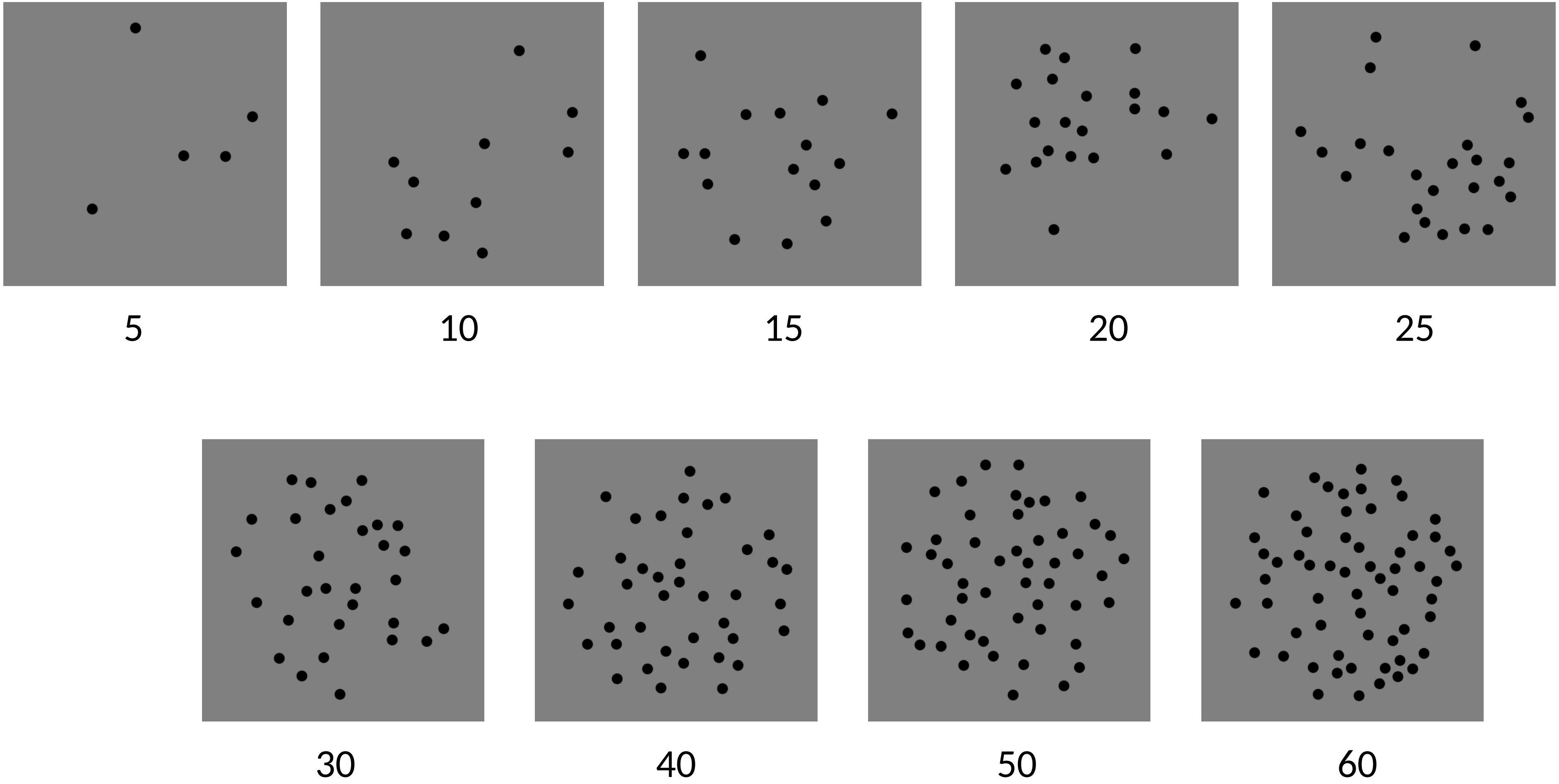
Perception of image degradation due to compression (Charrier et al. 2007)

Examples of MLDS usage

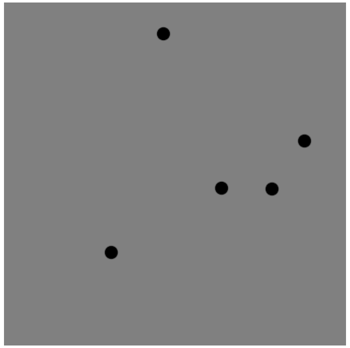
Perception of image degradation due to compression (Charrier et al. 2007)



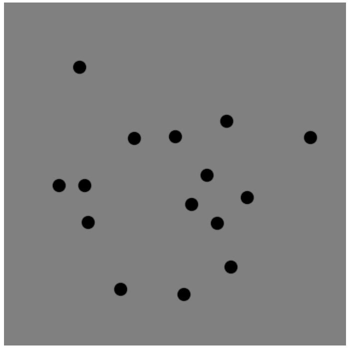
Tutorial today: perception of numerosity



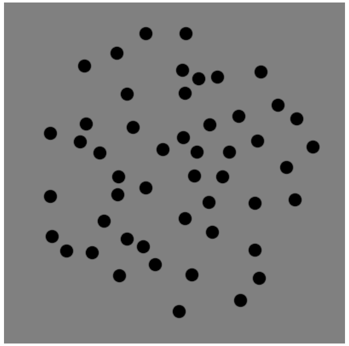
Tutorial today: perception of numerosity



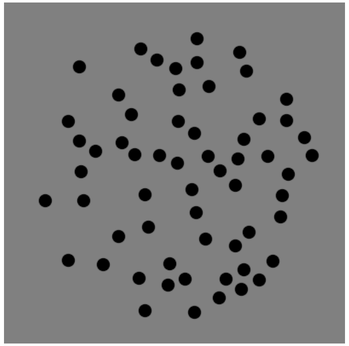
5



15



50



60

Exercise: acquire your own data

Go to

https://kutt.it/mlds-exp

and run the MLDS experiment with the method of triads

At the end of the experiment you will get a CSV file.

Keep it.

Observations?

- Mix of difficult and easy trials
- “Odd one out” type of judgment
- Did you use any strategy?

MLDS: details of experimental design

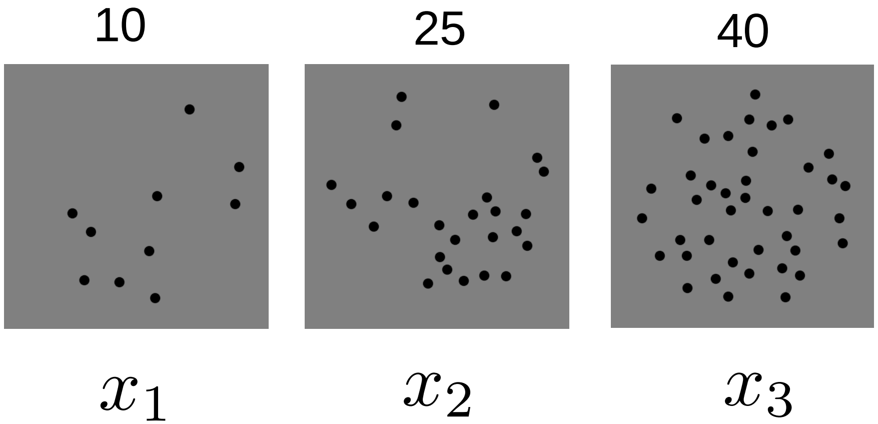
- Triads are always ascending $s_1 > s_2 > s_3$

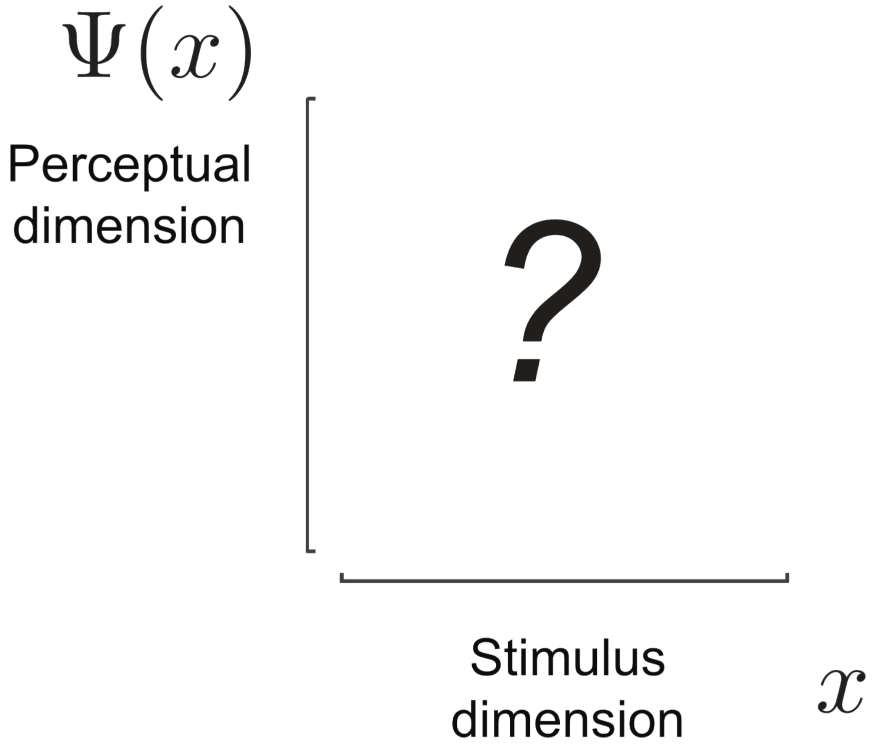
  or descending $s_1 < s_2 < s_3$

- Number of unique triads $\binom{N}{3}$ for N stimuli

- Repeat in block or random sampling

MLDS observer model and estimation





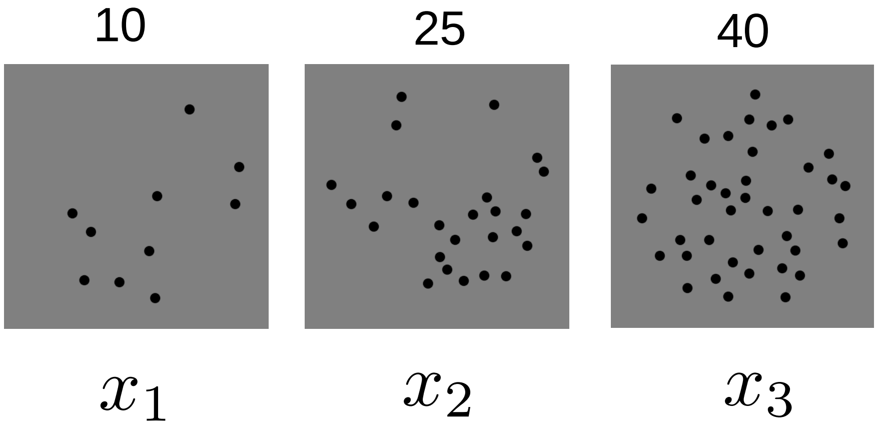
Decision model

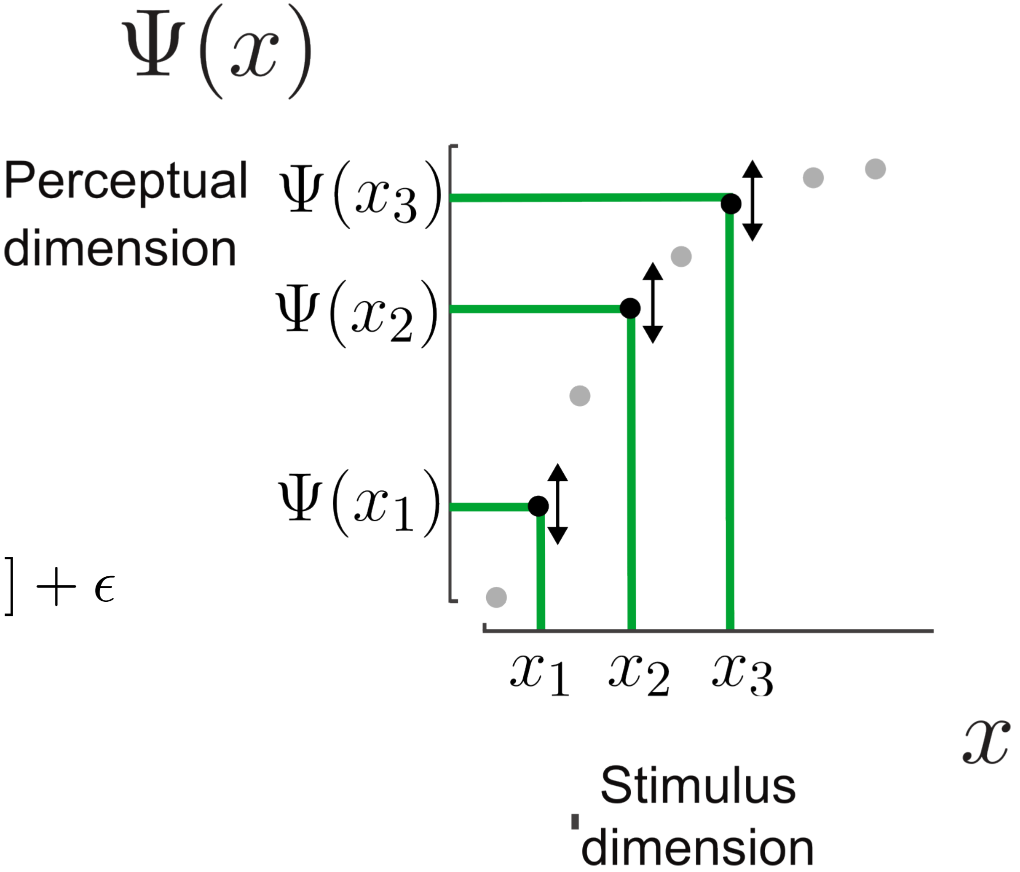
$$\Delta = [\Psi(x_3) - \Psi(x_2)] - [\Psi(x_2) - \Psi(x_1)] + \epsilon$$
$$= \Psi(x_3) - 2\Psi(x_2) + \Psi(x_1) + \epsilon$$

$$\epsilon \sim N(0, \sigma^2)$$

if $\Delta > 0 \rightarrow$ “right”
otherwise $\rightarrow$ “left”

MLDS observer model and estimation





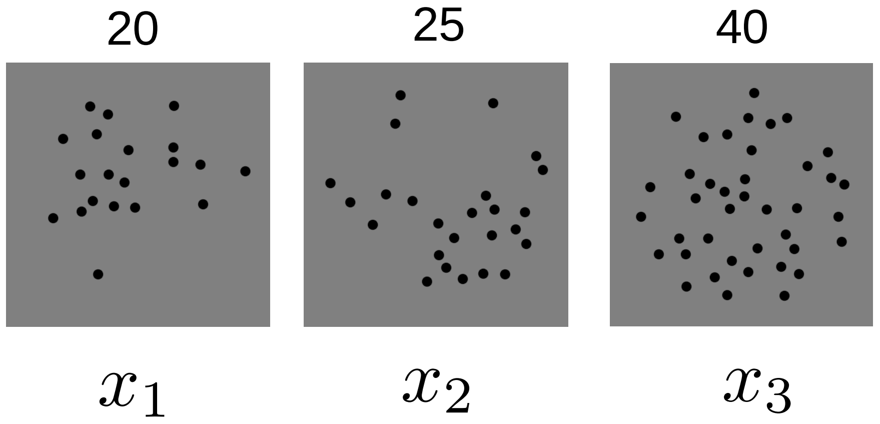
Decision model

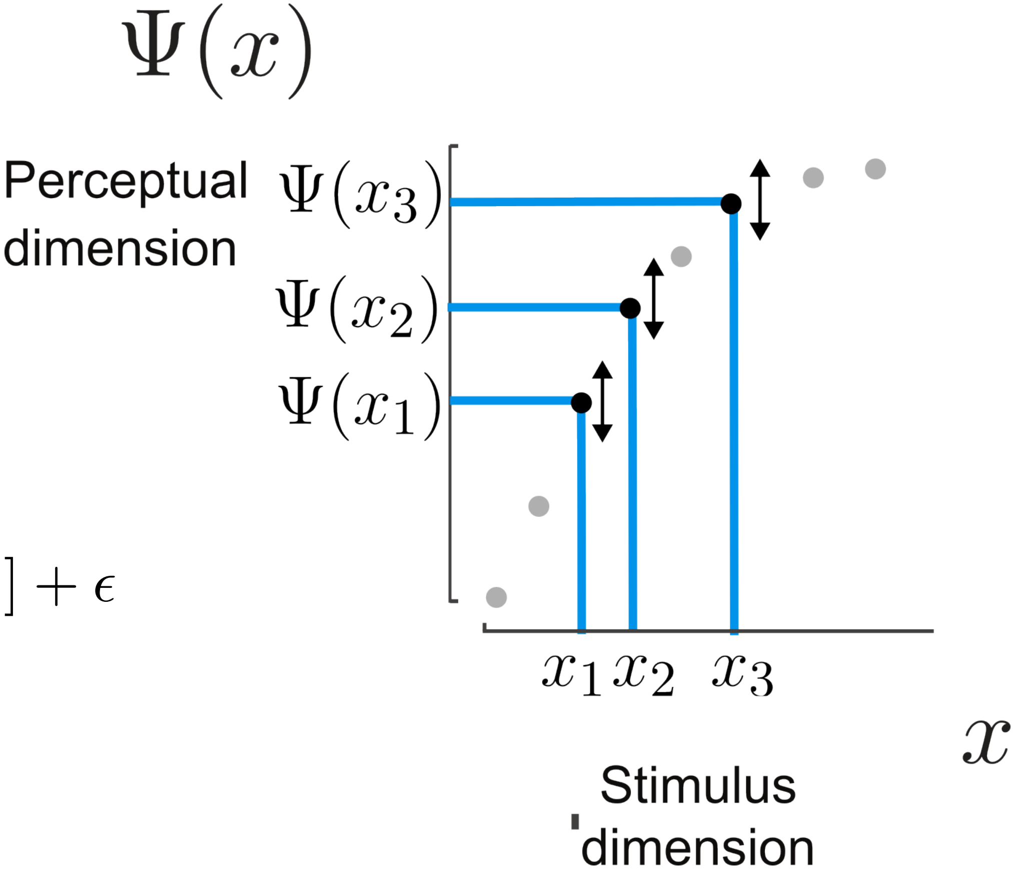
$$
\begin{aligned}
\Delta &= [\Psi(x_3) - \Psi(x_2)] - [\Psi(x_2) - \Psi(x_1)] + \epsilon \\
&= \Psi(x_3) - 2\Psi(x_2) + \Psi(x_1) + \epsilon
\end{aligned}
$$

$$\epsilon \sim N(0, \sigma^2)$$

$$
\begin{array}{ll}
\text{if} \quad \Delta > 0 & \rightarrow \text{“right”} \\
\text{otherwise} & \rightarrow \text{“left”}
\end{array}
$$

MLDS observer model and estimation





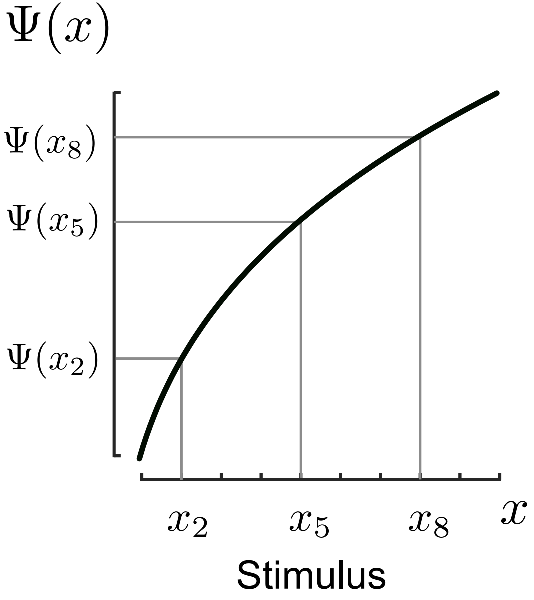
Decision model

$$\Delta = [\Psi(x_3) - \Psi(x_2)] - [\Psi(x_2) - \Psi(x_1)] + \epsilon$$
$$= \Psi(x_3) - 2\Psi(x_2) + \Psi(x_1) + \epsilon$$

$$\epsilon \sim N(0, \sigma^2)$$

if $\Delta > 0 \rightarrow$ "right"
otherwise $\rightarrow$ "left"

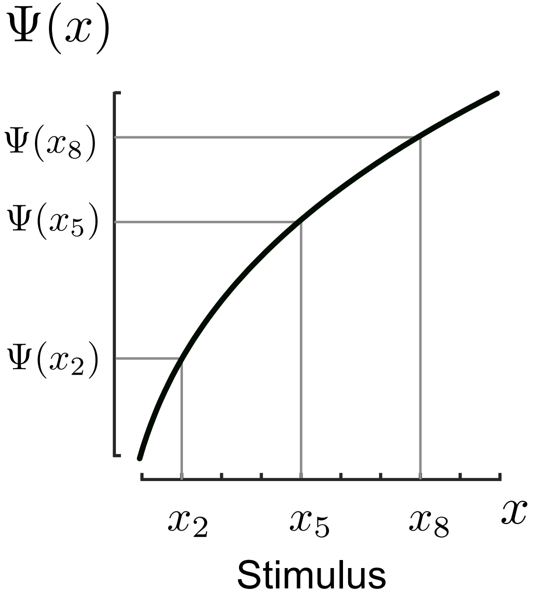
MLDS as a Generalized Linear Model (GLM)



Decision model $$\Delta = \Psi(x_8) - 2\Psi(x_5) + \Psi(x_2)$$

Wiebel, Aguilar & Maertens (2017)

MLDS as a Generalized Linear Model (GLM)



$$g(\mathrm{E}[\boldsymbol{Y}]) = \boldsymbol{X} \ * \boldsymbol{\beta} \ + \ \epsilon$$

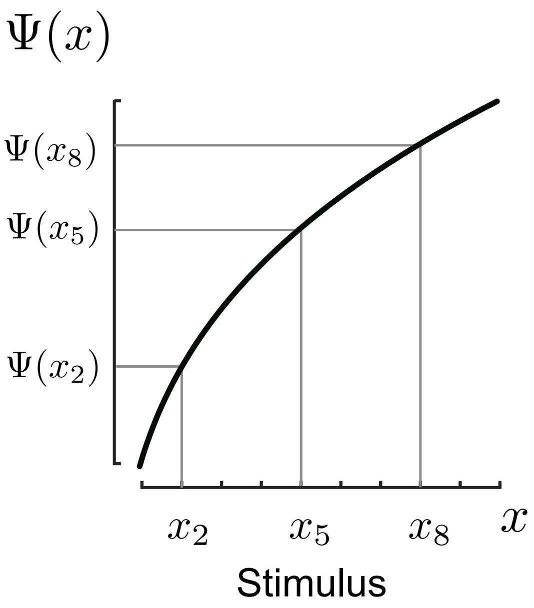
$$\begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ \ \\ \vdots \end{bmatrix} \quad \begin{bmatrix} 0 & 1 & 0 & 0 & -2 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & -2 & 1 & 0 \\ 0 & 1 & 0 & 0 & -2 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & -2 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & -2 & 0 & 0 & 0 & 1 \\ & & & & \vdots & & & & & \end{bmatrix} \quad \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \\ \beta_6 \\ \beta_7 \\ \beta_8 \\ \beta_9 \\ \beta_{10} \end{bmatrix}$$

Decision model $$\Delta = \Psi(x_8) - 2\Psi(x_5) + \Psi(x_2)$$

GLM $$g(E[Y]) = \mathbf{1} \cdot \beta_8 - \mathbf{2} \cdot \beta_5 + \mathbf{1} \cdot \beta_2$$

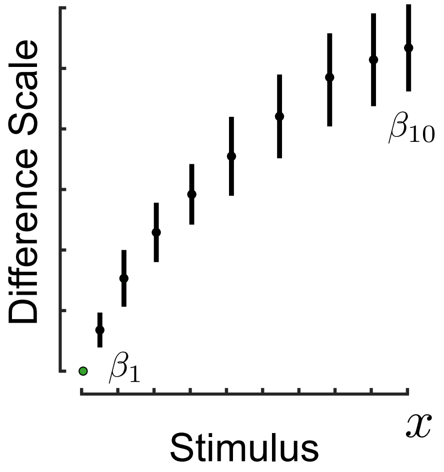
Wiebel, Aguilar & Maertens (2017)

MLDS as a Generalized Linear Model (GLM)



$$g(\mathrm{E}[\boldsymbol{Y}]) = \boldsymbol{X} \ * \boldsymbol{\beta} + \ \epsilon$$

$$\begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ \\ \vdots \end{bmatrix} \quad \begin{bmatrix} 0 & 1 & 0 & 0 & -2 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & -2 & 1 & 0 \\ 0 & 1 & 0 & 0 & -2 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & -2 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & -2 & 0 & 0 & 0 & 1 \\ & & & & \vdots & & & & & \end{bmatrix} \quad \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \\ \beta_6 \\ \beta_7 \\ \beta_8 \\ \beta_9 \\ \beta_{10} \end{bmatrix}$$



Decision model $$\Delta = \Psi(x_8) - 2\Psi(x_5) + \Psi(x_2)$$

GLM $$g(E[Y]) = \mathbf{1} \cdot \beta_8 - \mathbf{2} \cdot \beta_5 + \mathbf{1} \cdot \beta_2$$

Wiebel, Aguilar & Maertens (2017)

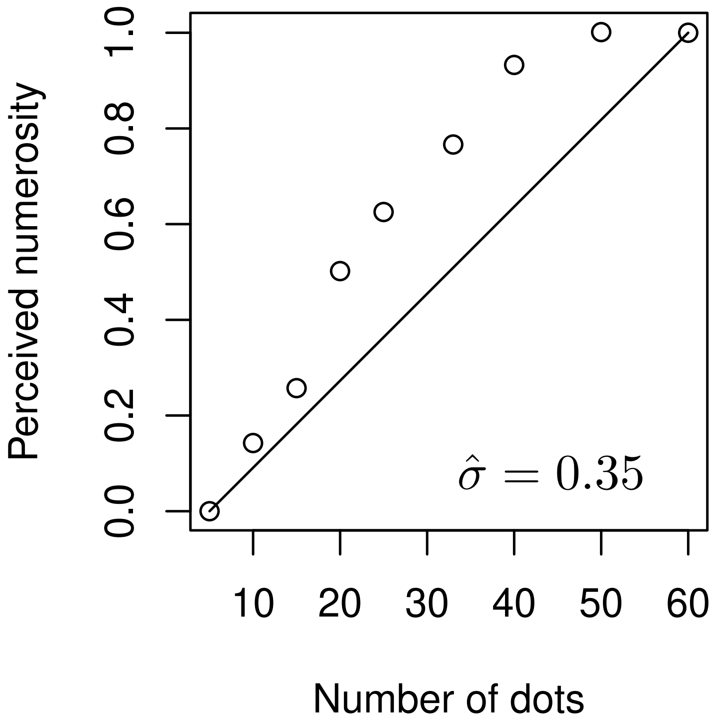
Live coding:

analyzing MLDS data in R

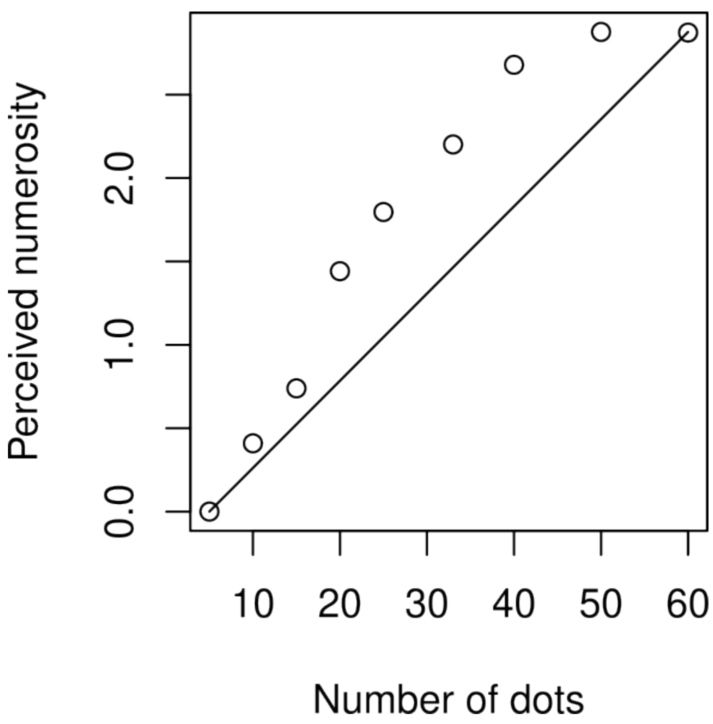
Interpretation of results

Two parametrizations

"standard": range from 0 to 1



"unconstrained": range from 0 to $\frac{1}{\hat{\sigma}}$



Exercise: analyze your own data

UPDATE

1) Got to github and download the repository

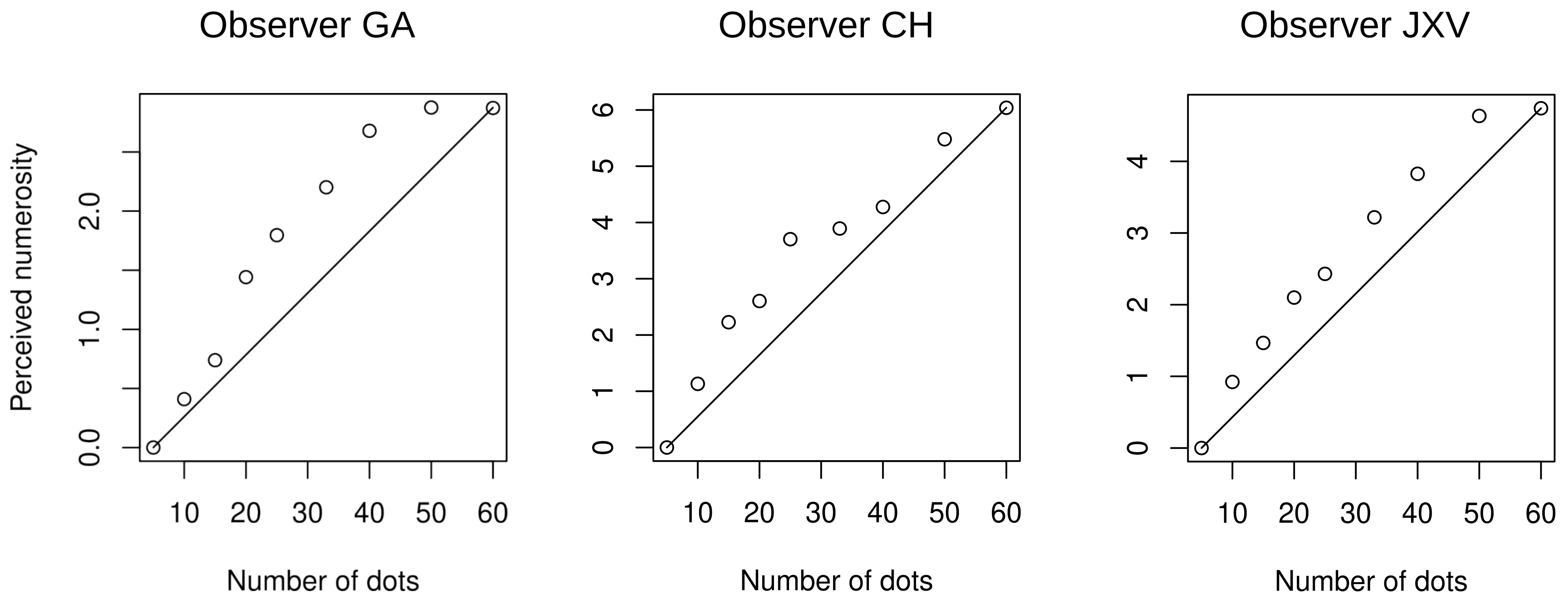
   https://github.com/computational-psychology/ecvp2025-mlds-mlcm-tutorial

2) Copy your CSV files to the folder data/mlds/

3) In Rstudio, open file: tutorial_mlds.Rmd

4) Edit lines XX to match the filename of your data

5) Run all cells (top right, *Run / Run all*)


Observer GA
Observer CH
Observer JXV
Perceived numerosity
0.0
1.0
2.0
0
1
2
3
4
5
6
10
20
30
40
50
60
Number of dots
Number of dots
Number of dots

Confidence intervals

GLM

parameters ≈ perceptual scale

Experimental data $\longrightarrow$ $\boxed{g(E[\boldsymbol{Y}]) = X\boldsymbol{\beta} + \epsilon}$ $\longrightarrow$ $\hat{\boldsymbol{\beta}}$

Confidence intervals

GLM

Experimental data → $$g(E[\mathbf{Y}]) = X\boldsymbol{\beta} + \epsilon$$ → $\hat{\boldsymbol{\beta}}$

parameters ≈ perceptual scale

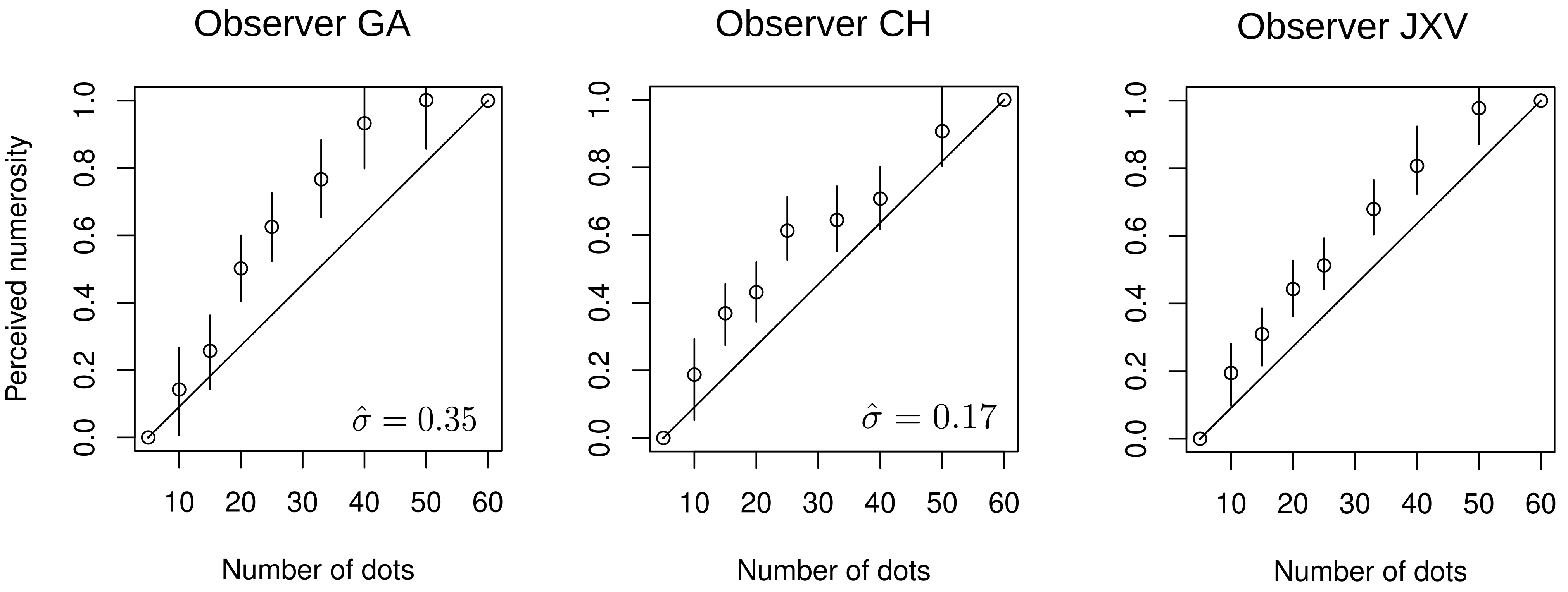
$\hat{\boldsymbol{p}}$ ← $$\hat{\boldsymbol{p}} = g^{-1}(X\hat{\boldsymbol{\beta}})$$

simulate exp
x 10000

Surrogate data → $$g(E[\mathbf{Y}]) = X\boldsymbol{\beta} + \epsilon$$ → $\hat{\boldsymbol{\beta}}^*$ → Confidence intervals

Distribution
bootstrapped
parameters

Confidence intervals



To keep in mind - Model assumptions

- observer is performing the difference as stated in the decision model

- observer is able to sort stimuli (weak order assumption, Krantz et al.)

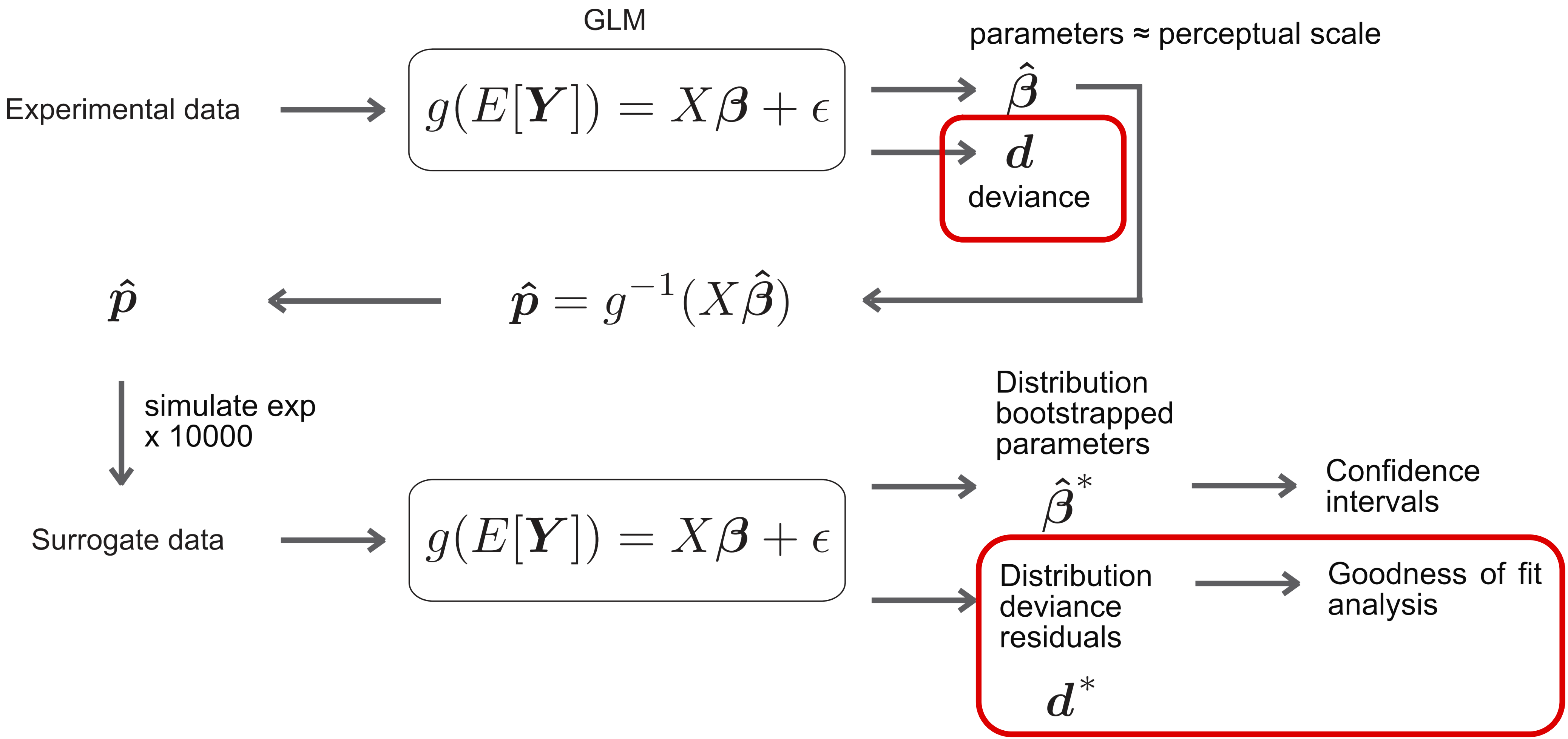
- i.i.d. assumptions: each trial is a Bernoulli variable, independent of each other (e.g. no learning)

We can test the integrity of assumptions via simulation

Logic

if the observer model holds,
then experimental data is no different than simulated data created using the observer model

Testing model assumptions: goodness of fit evaluation



Goodness of fit evaluation - plots

Goodness of fit evaluation - plots

Check 1: (deviance) residuals

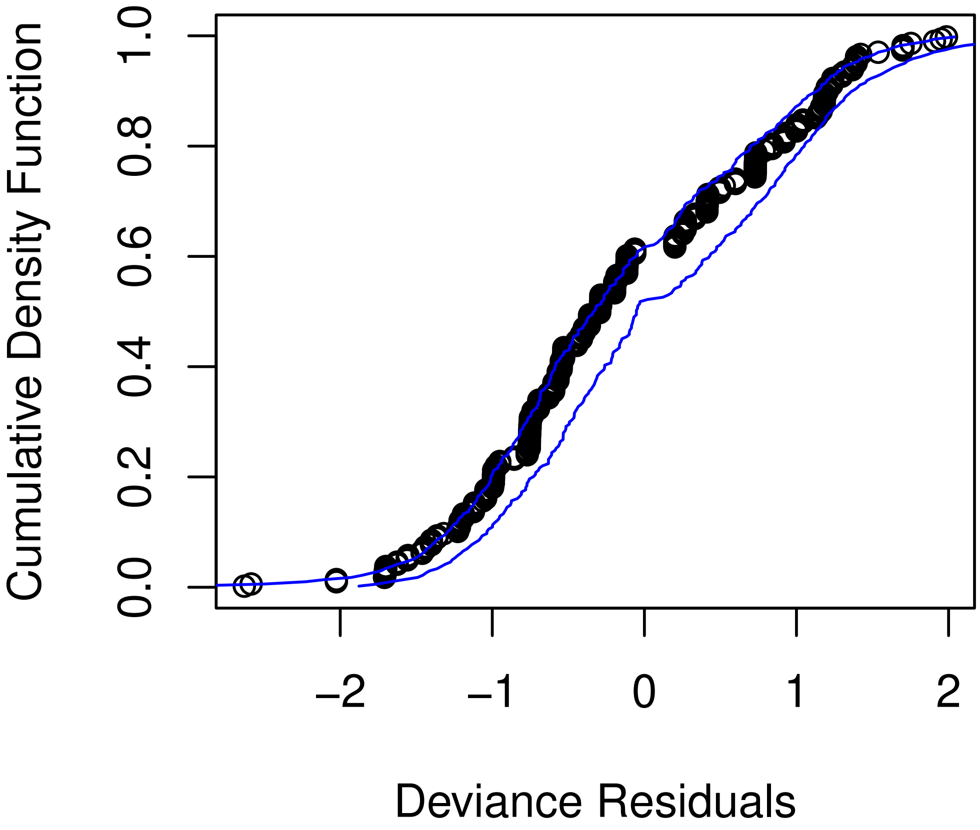
- cumulative function of deviance residuals (markers)

- 95% envelope from *simulated* deviance residuals (blue lines)

This envelope represents *expected* range

For an appropriate goodness of fit:

data (markers) should be inside the range



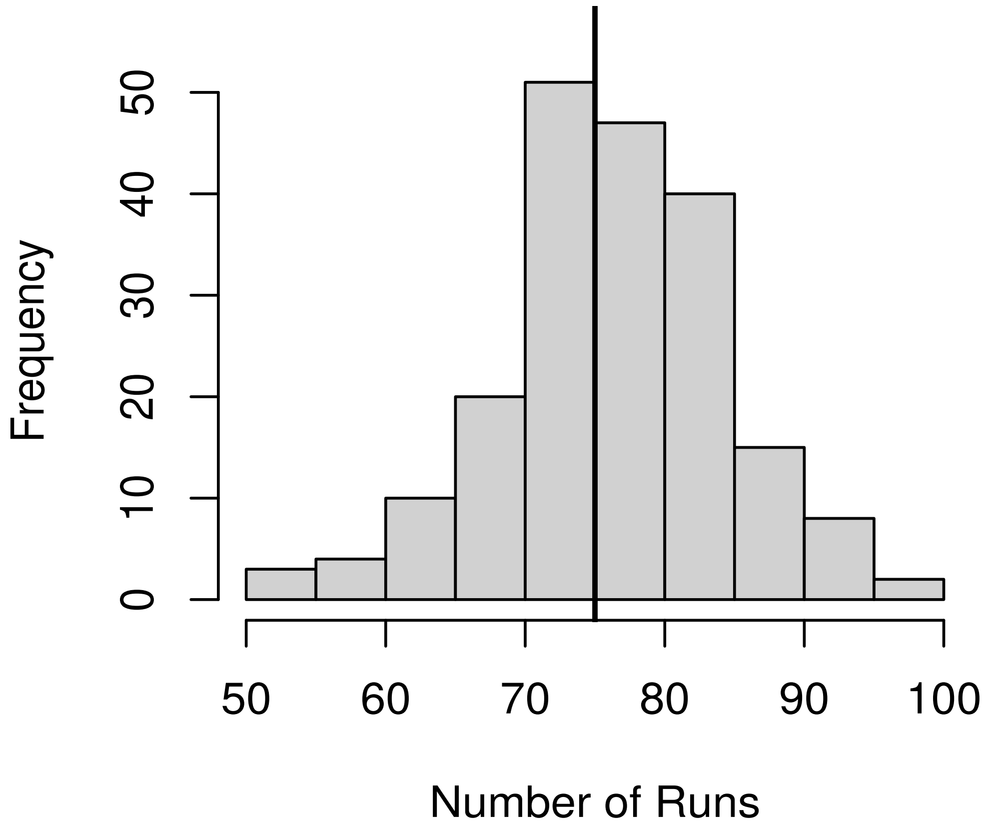
Goodness of fit evaluation - plots

Check 2: distribution of deviance crossings or "runs"

- sort deviance residuals by linear predictor

- count how many have consecutively the same sign (vertical line)

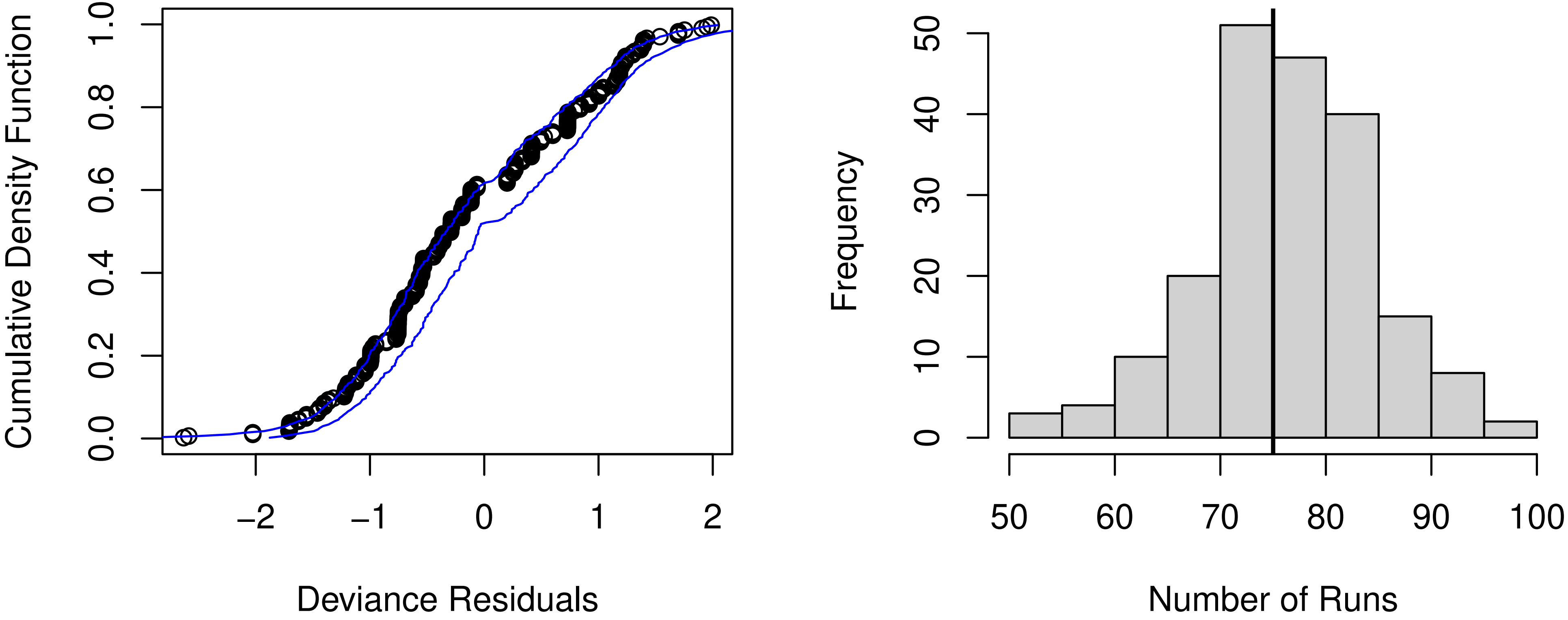
- do the same for the *simulated* deviance residuals, and plot as histogram



For an appropriate goodness of fit:

"Runs" from data (vertical line) should not be sig. different than the distribution

Goodness of fit evaluation - plots

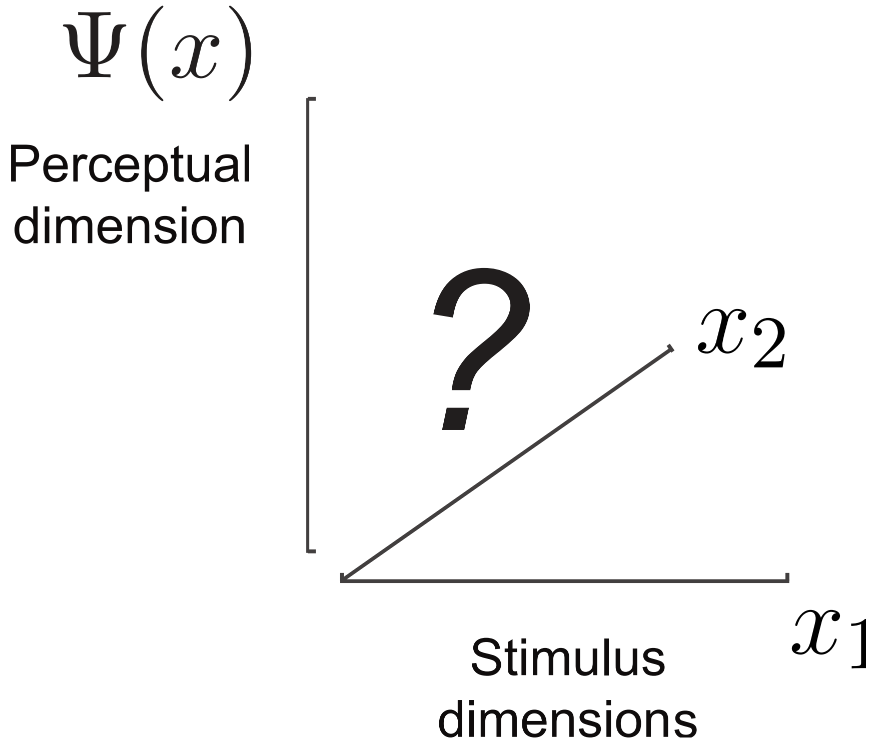


Maximum Likelihood Conjoint Measurement (MLCM)

Motivation for conjoint measurement

What happens when more than 1 stimulus dimensions affect how stimuli look like?

→ conjoint measurement



Motivation for conjoint measurement

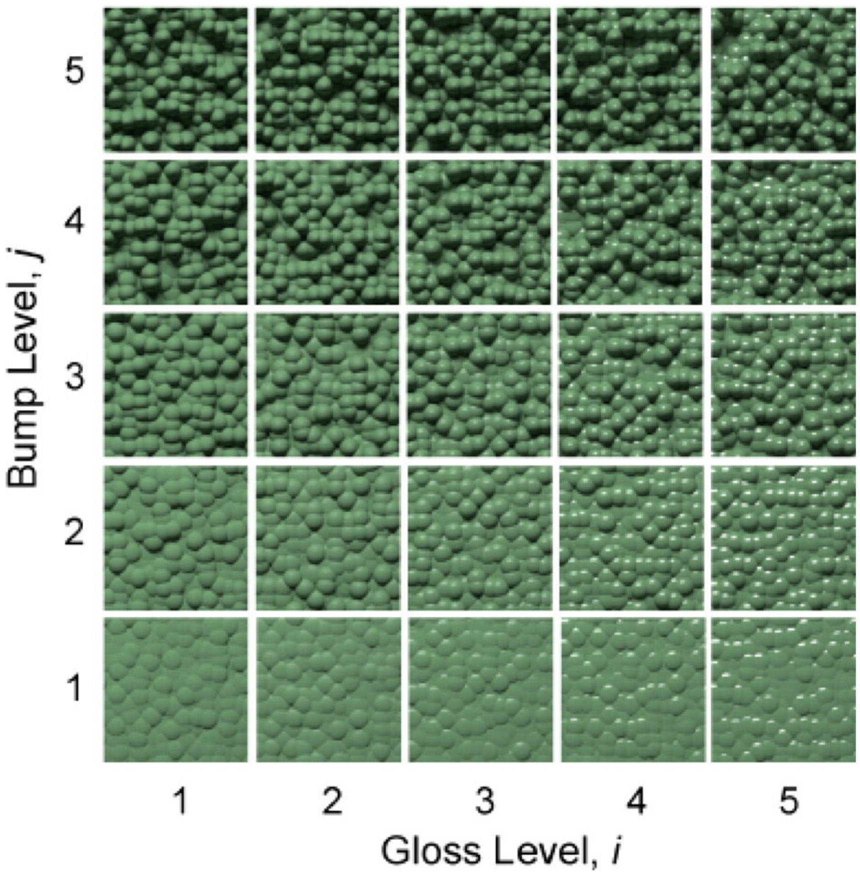
What happens when more than 1 stimulus dimensions affect how stimuli look like?

→ conjoint measurement

Example:

Perceived bumpiness (depth) depends on
- physical bumpiness (depth) and
- gloss

(Ho et al. 2008)



Motivation for conjoint measurement

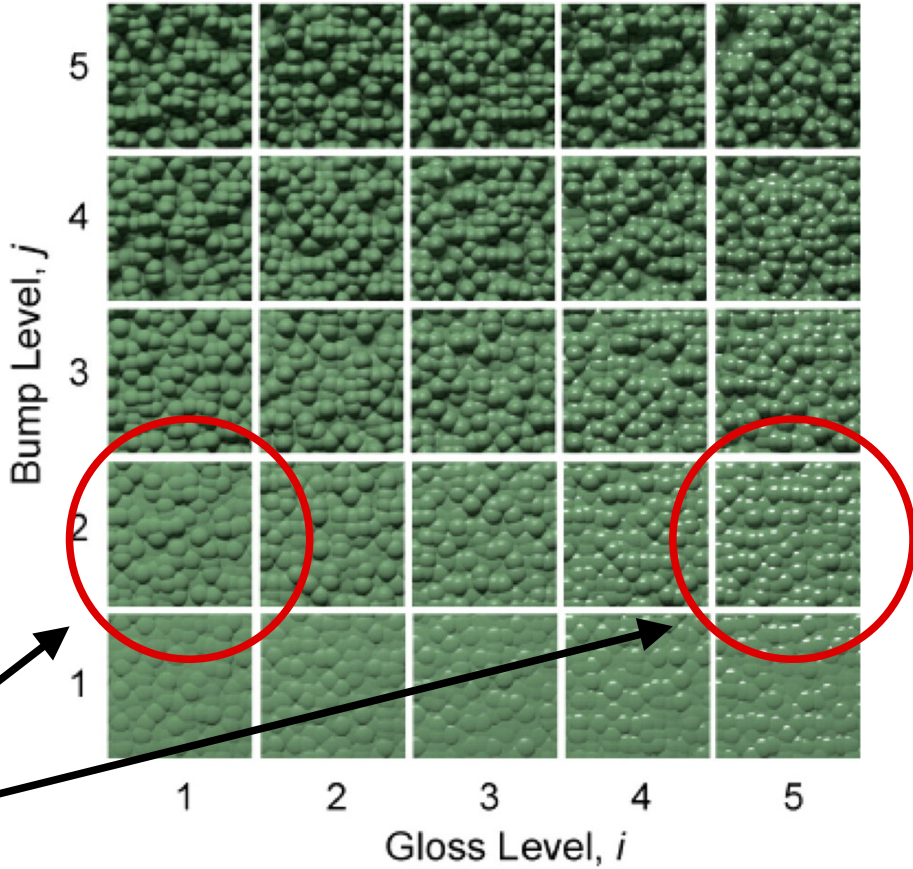
What happens when more than 1 stimulus dimensions affect how stimuli look like?

→ conjoint measurement

Example:

Perceived bumpiness (depth) depends on
- physical bumpiness (depth) and
- gloss

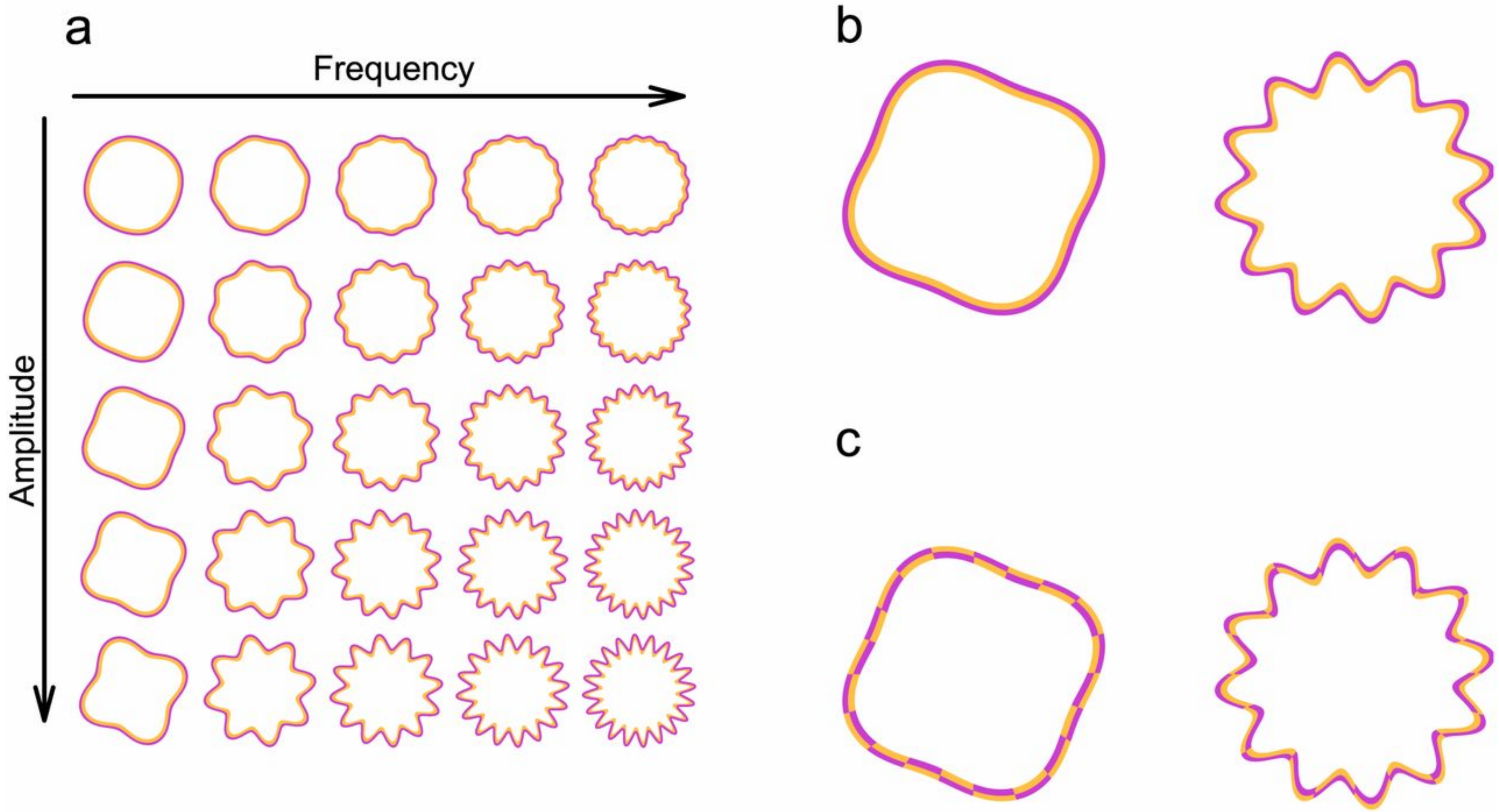
(Ho et al. 2008)



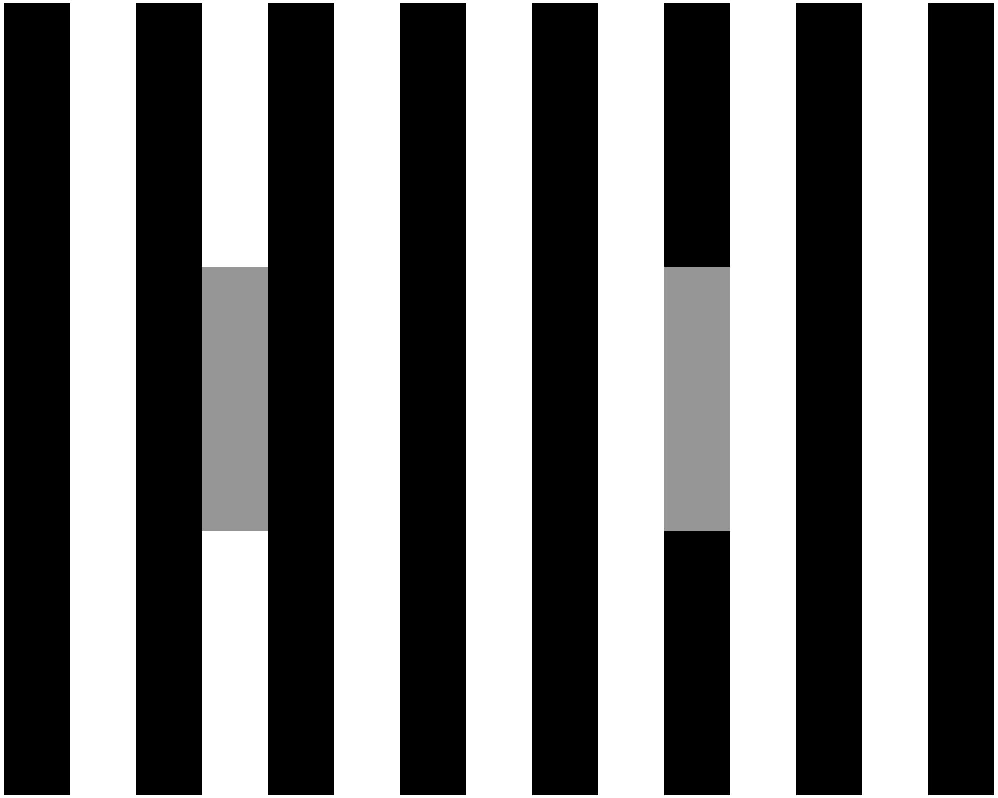
Same physical "bumpiness", glossier stimulus looks more bumpy than matte one

Examples of MLCM usage

Perception of the watercolor effect (Gerardin et al. 2014)

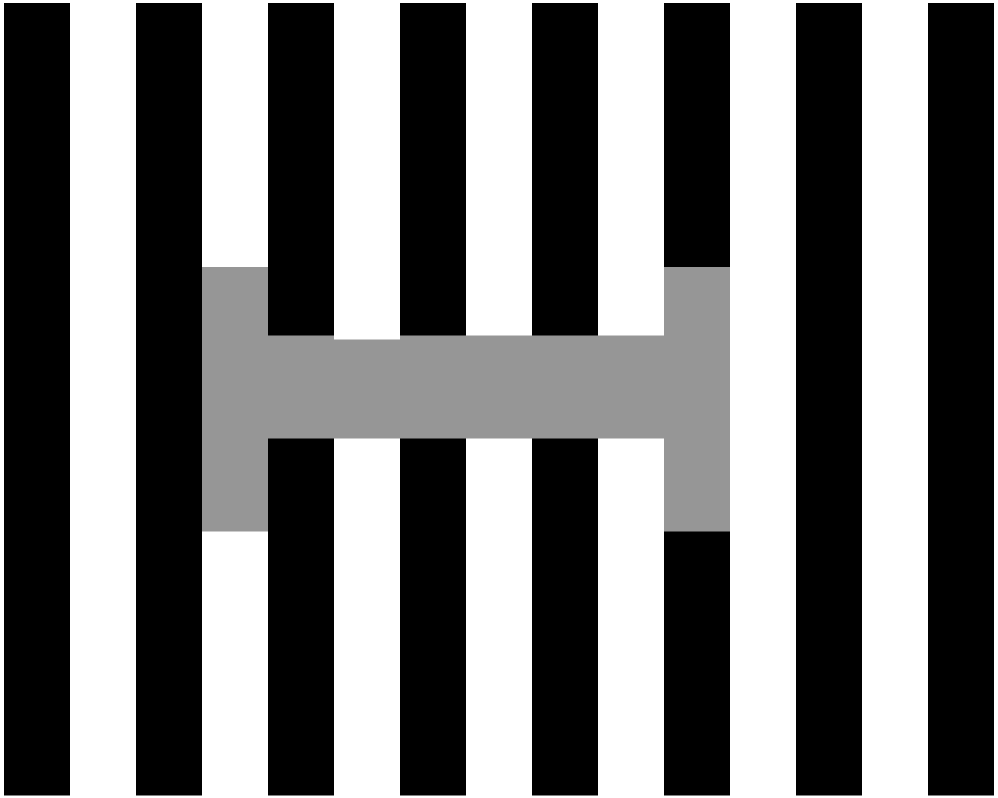


Example today: brightness in White's illusion

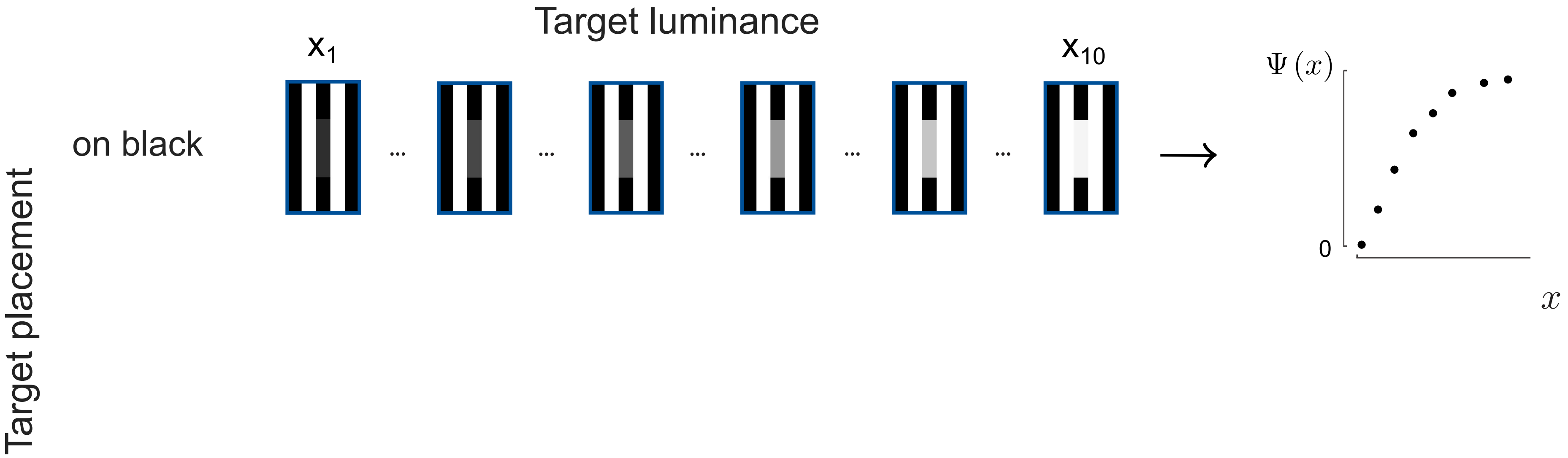


White (1979)

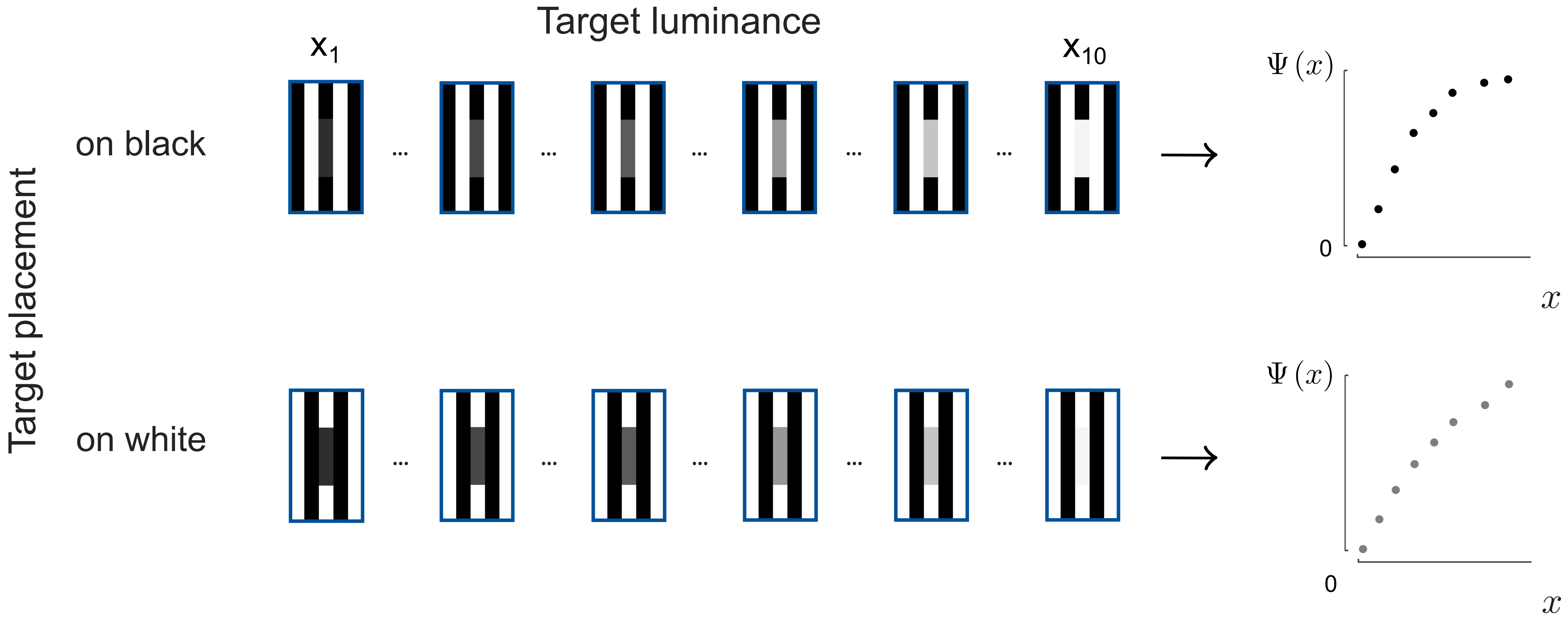
Example today: brightness in White's illusion



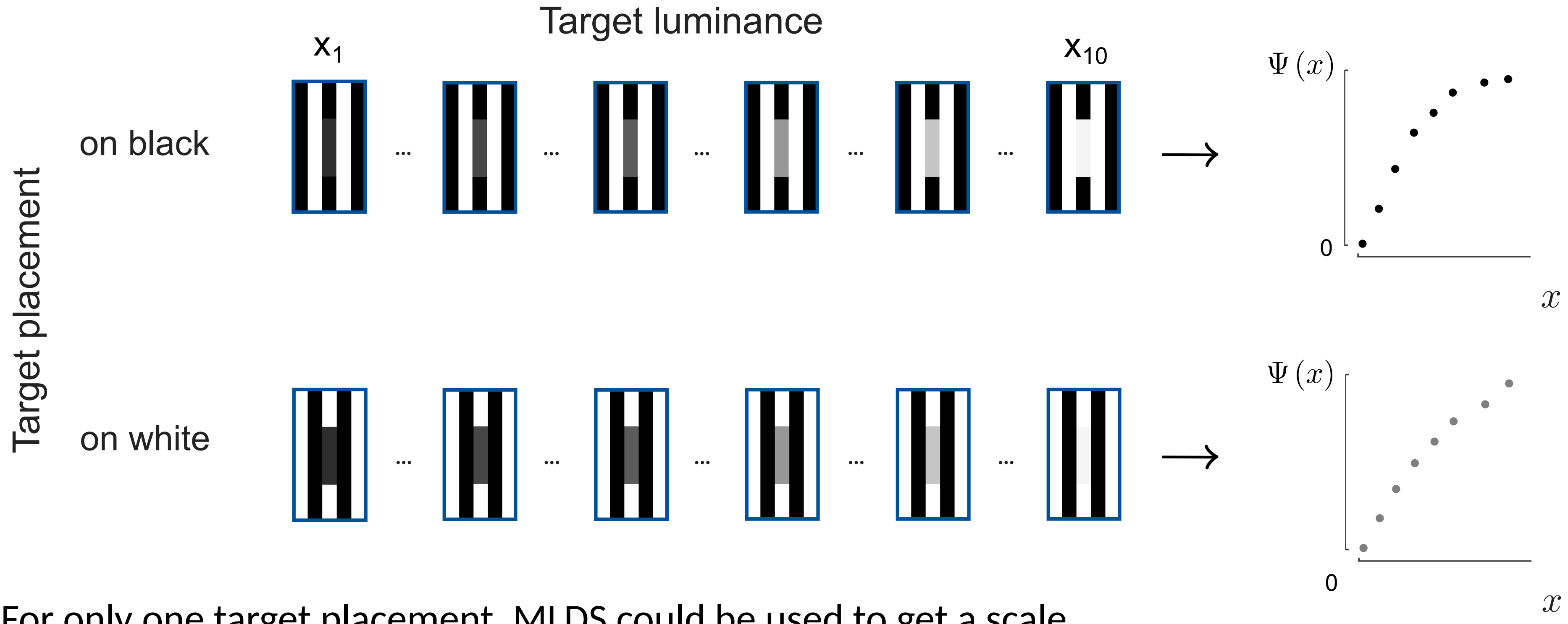
White (1979)



Target luminance
Target placement
on black
x1
x10
...
Ψ (x)
0
x



Vincent, Maertens & Aguilar (2024)



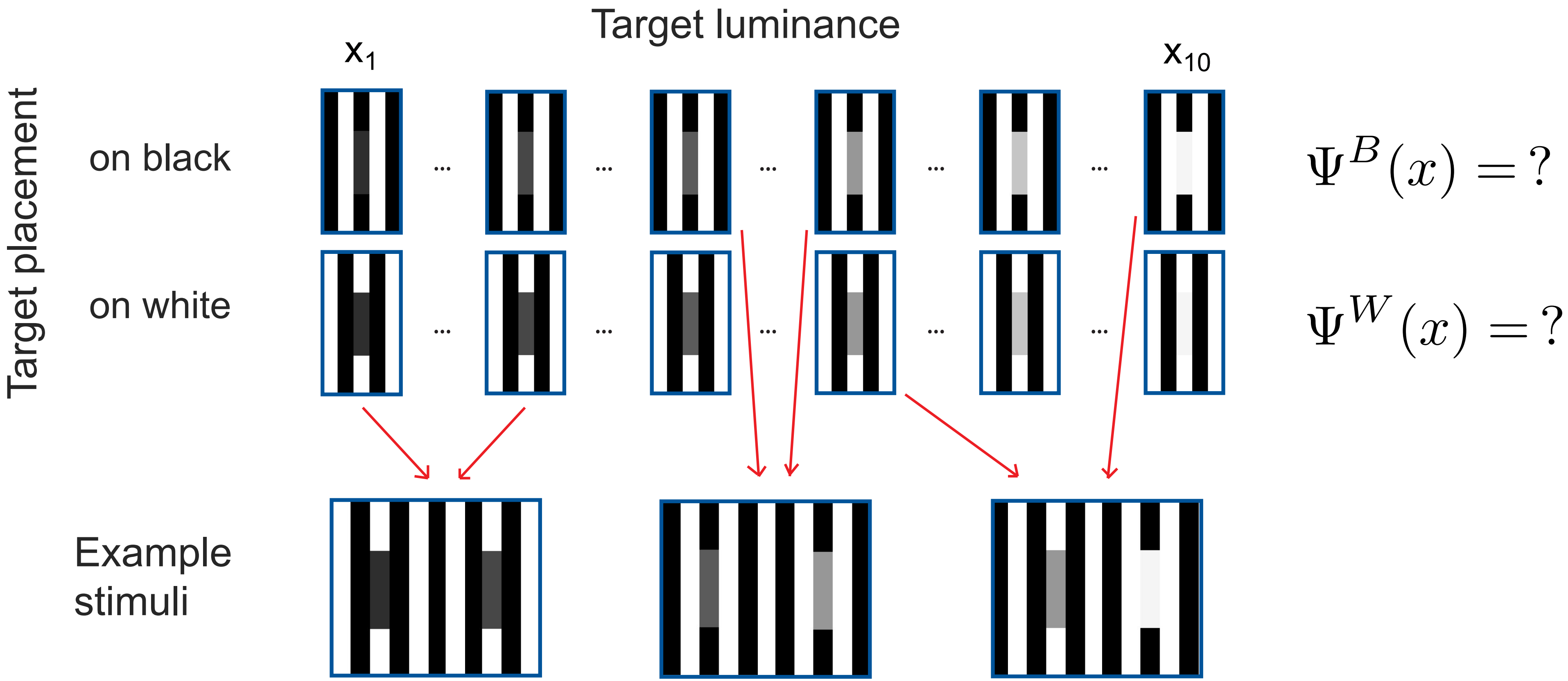
For only one target placement, MLDS could be used to get a scale

BUT! We would obtain two independent scales, both anchored at zero.

Scales can be scaled and shifted, so we can't relate them meaningfully to each other

We need to add **target placement** as a **2nd stimulus dimension**
→ MLCM

Vincent, Maertens & Aguilar (2024)



Vincent, Maertens & Aguilar (2024)

Experimental design in MLCM

Stimuli vary in **two or more stimulus dimensions**

Dim 2

Dim 1

$s_{(1,1)}$	$s_{(1,2)}$	$s_{(1,3)}$	...	$s_{(1,n)}$
$s_{(2,1)}$	$s_{(2,2)}$	$s_{(2,3)}$		
$s_{(3,1)}$	$s_{(3,2)}$	...		
...			...	
$s_{(m,1)}$				...

Paired comparisons

$$x_1 \longleftrightarrow x_2$$

Which stimulus is

brighter?
glossier?
...
<insert your perceptual dimension> ?

Experimental design in MLCM

Stimuli vary in **two or more stimulus dimensions**

Dim 2

Dim 1

$s_{(1,1)}$	$s_{(1,2)}$	$s_{(1,3)}$	...	$s_{(1,n)}$
$s_{(2,1)}$	$s_{(2,2)}$	$s_{(2,3)}$		
$s_{(3,1)}$	$s_{(3,2)}$	...		
...			...	
$s_{(m,1)}$				...

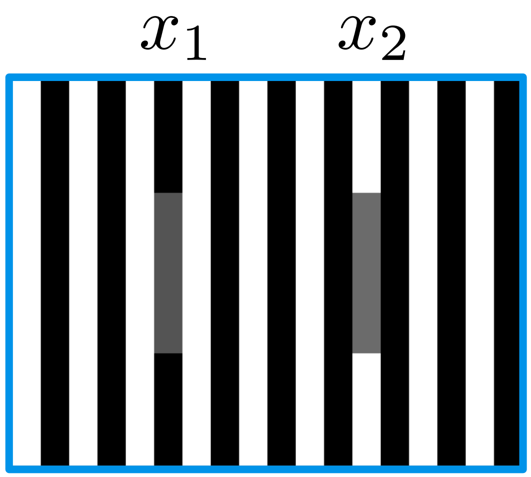
Paired comparisons

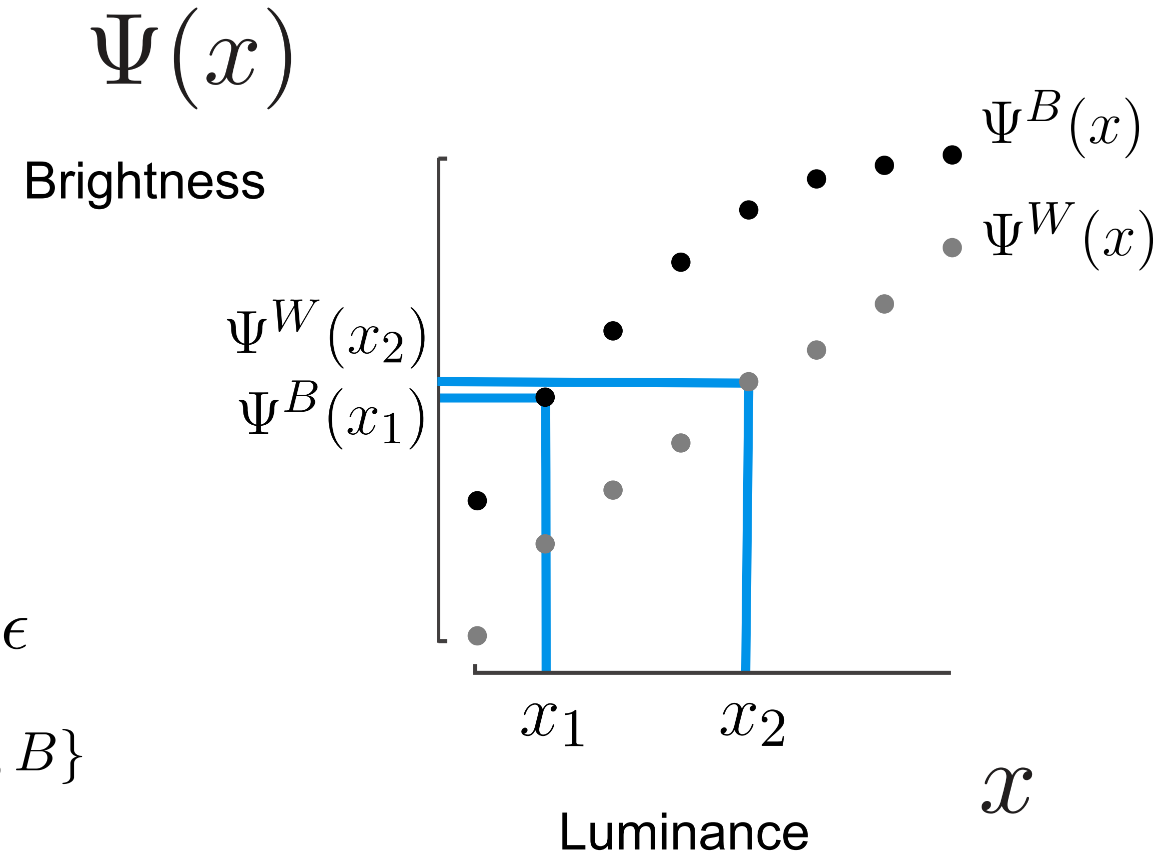
$$x_1 \longleftrightarrow x_2$$

Which stimulus is

brighter?
glossier?
...
<insert your perceptual dimension> ?

MLCM observer model(s)





Decision model

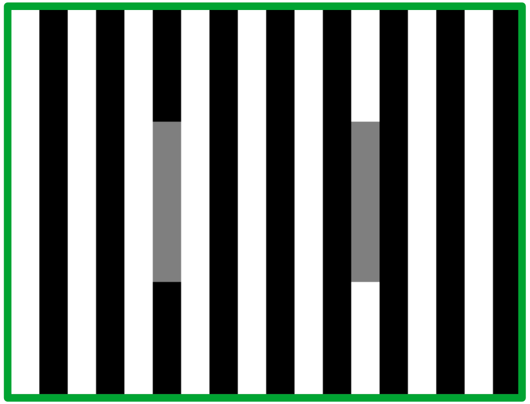
$$\Delta = \Psi^{C_2}(x_2) - \Psi^{C_1}(x_1) + \epsilon$$

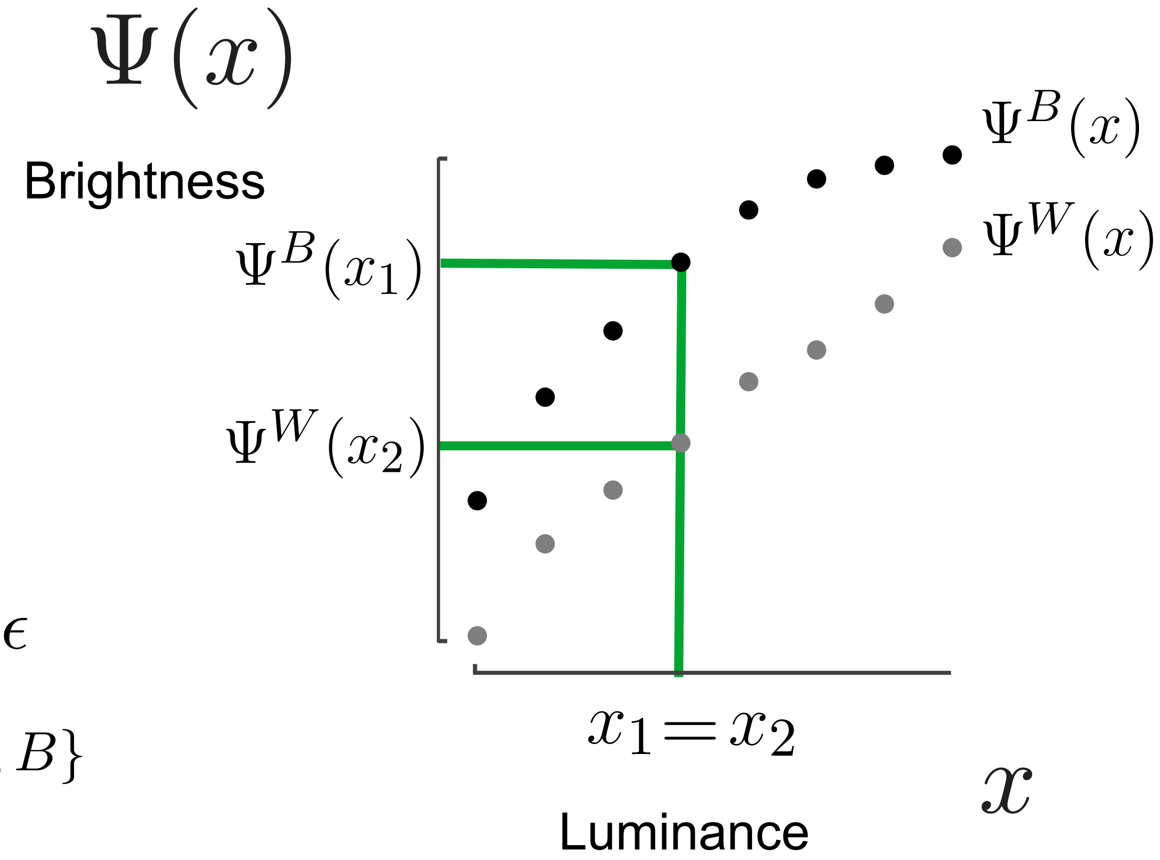
$$\epsilon \sim N(0, \sigma^2) \qquad C_1, C_2 \in \{W, B\}$$

if $\Delta > 0 \rightarrow$ "right"
otherwise $\rightarrow$ "left"

MLCM observer model(s)

$x_1 = x_2$





Decision model

$$\Delta = \Psi^{C_2}(x_2) - \Psi^{C_1}(x_1) + \epsilon$$

$$\epsilon \sim N(0, \sigma^2) \qquad C_1, C_2 \in \{W, B\}$$

$$\begin{array}{ll} \text{if} \quad \Delta > 0 & \rightarrow \text{“right”} \\ \text{otherwise} & \rightarrow \text{“left”} \end{array}$$

Three observer models in MLCM

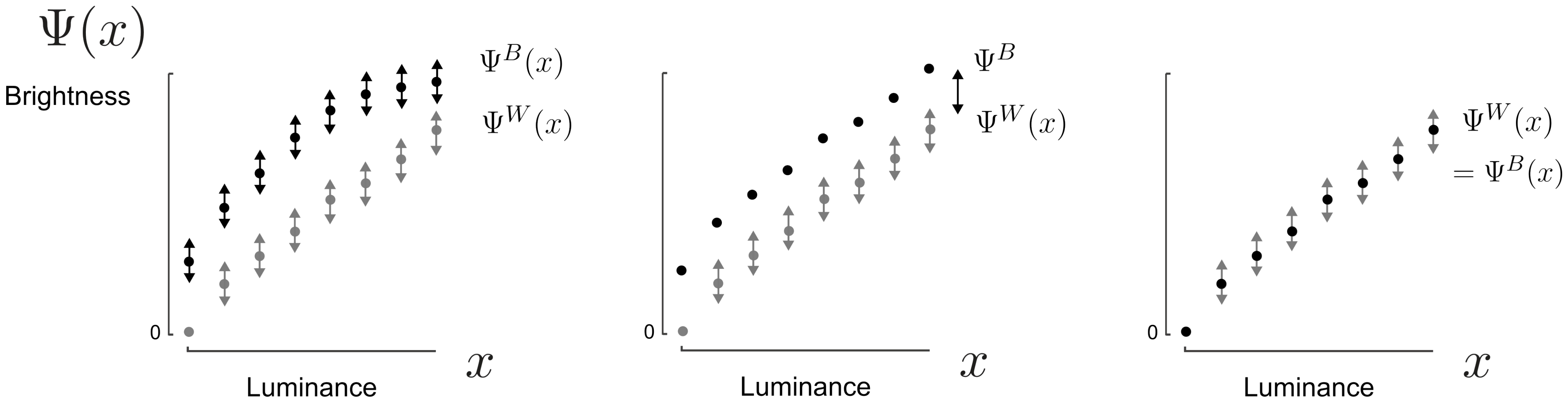
Full
(also called “saturated”)
All scale values fitted

Additive
Values for one dimension fitted, effect of other dimension is additive

Independent

Dimensions are Independent

More general → More constrained



→ Need of statistical model comparison

Exercise: analyze MLCM data

Run the MLCM experiment on White's illusion (Optional)

https://kutt.it/mlcm-exp

Session 1: all comparisons
Session 2: only comparisons across contexts

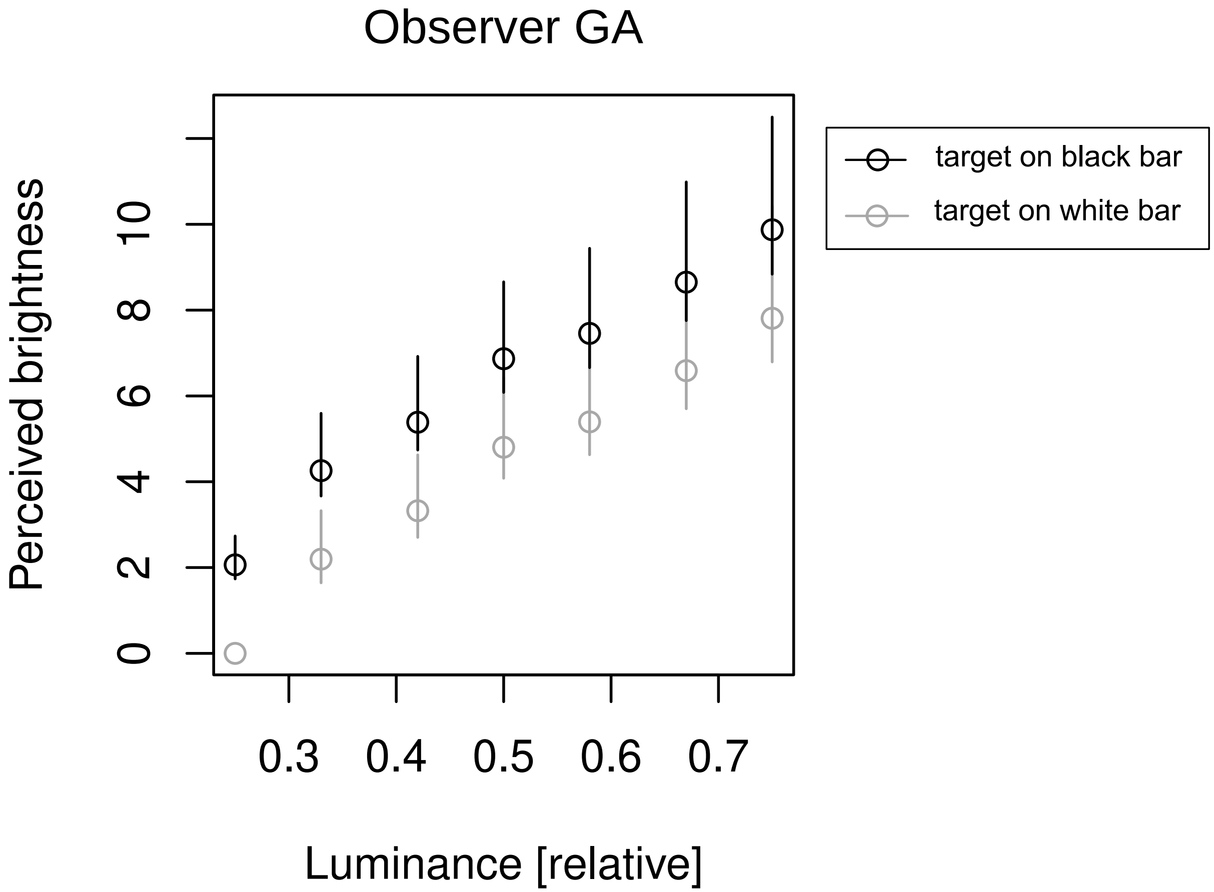
Copy your CSV files to the folder data/mlcm/

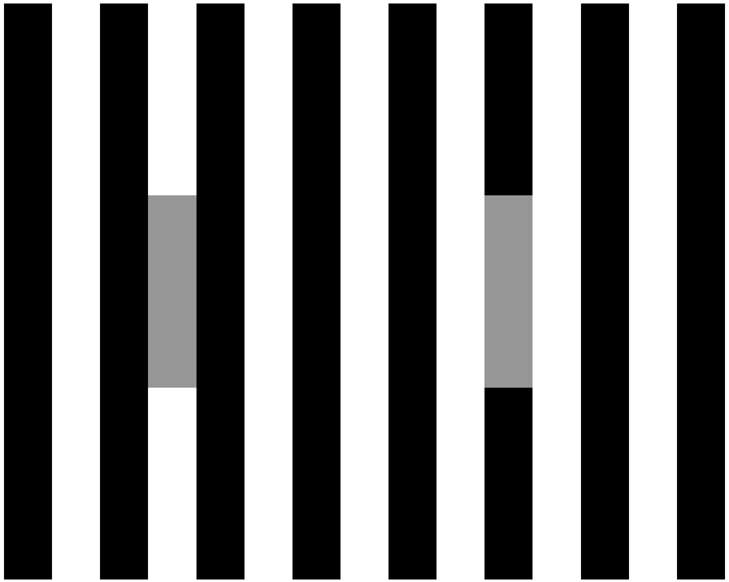
Analyse MLCM data

In Rstudio, open file: tutorial_mlcm.Rmd

Run all cells (top right, *Run / Run all*)

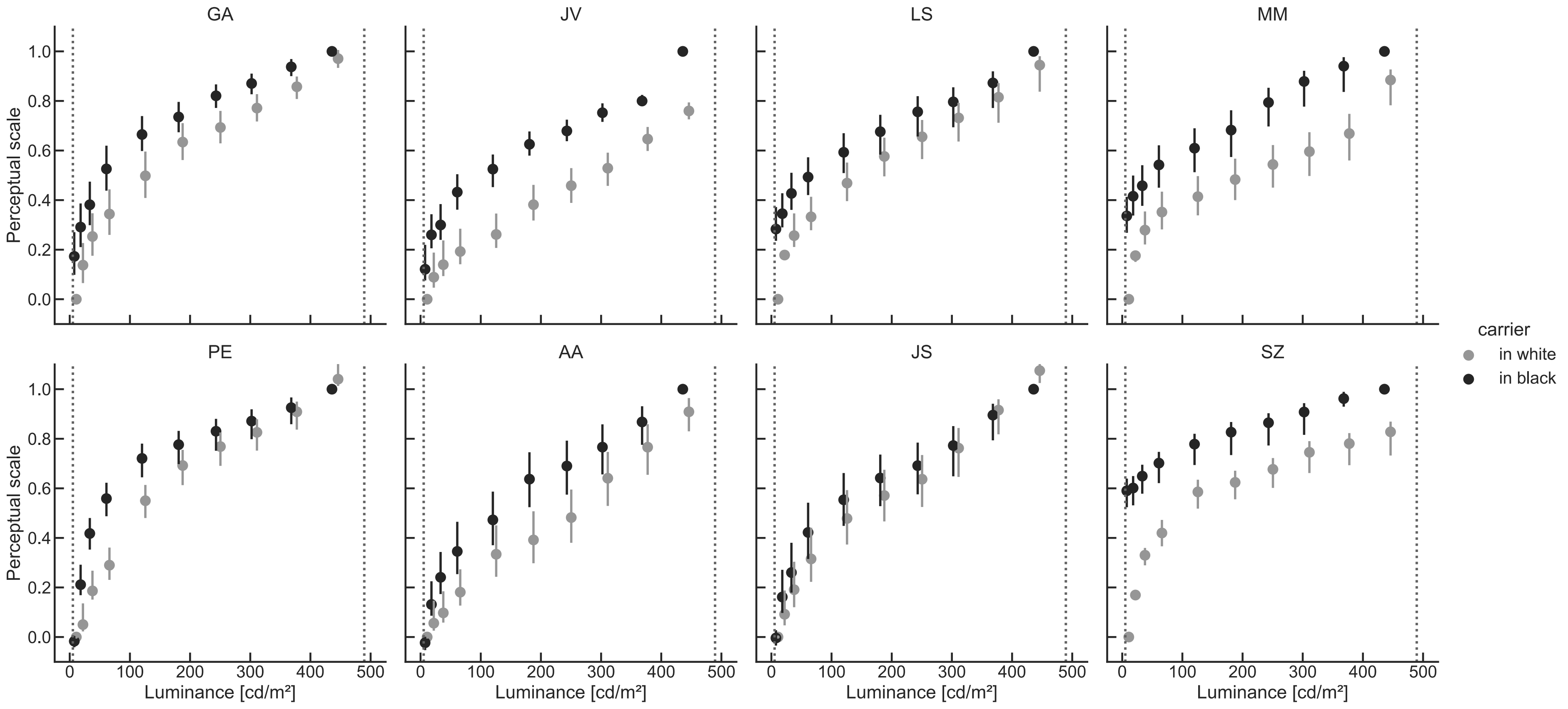
Interpretation of results





Interpretation of results

Scales for White’s illusion (Vincent, Maertens & Aguilar 2024)



Confidence intervals and goodness of fit evaluation

Same as for MLDS!

- similar functions in R

  boot.mlds() → *boot.mlcm()*

- same *binom.diagnostics()* for GoF evaluation

- same logic and interpretation

To keep in mind – Practical considerations

Choice of stimulus spacing

→ one should be able to order the stimuli (perceptually)

→ that means spacing should be supra-threshold

→ but not only easy trials, both methods require “hard” comparisons, for which the choice probability is different than 0 or 1 (always pressing only one option)

→ this sometimes requires piloting

Repeats: fixed repeat number vs. random sampling

Summary – MLDS and MLCM

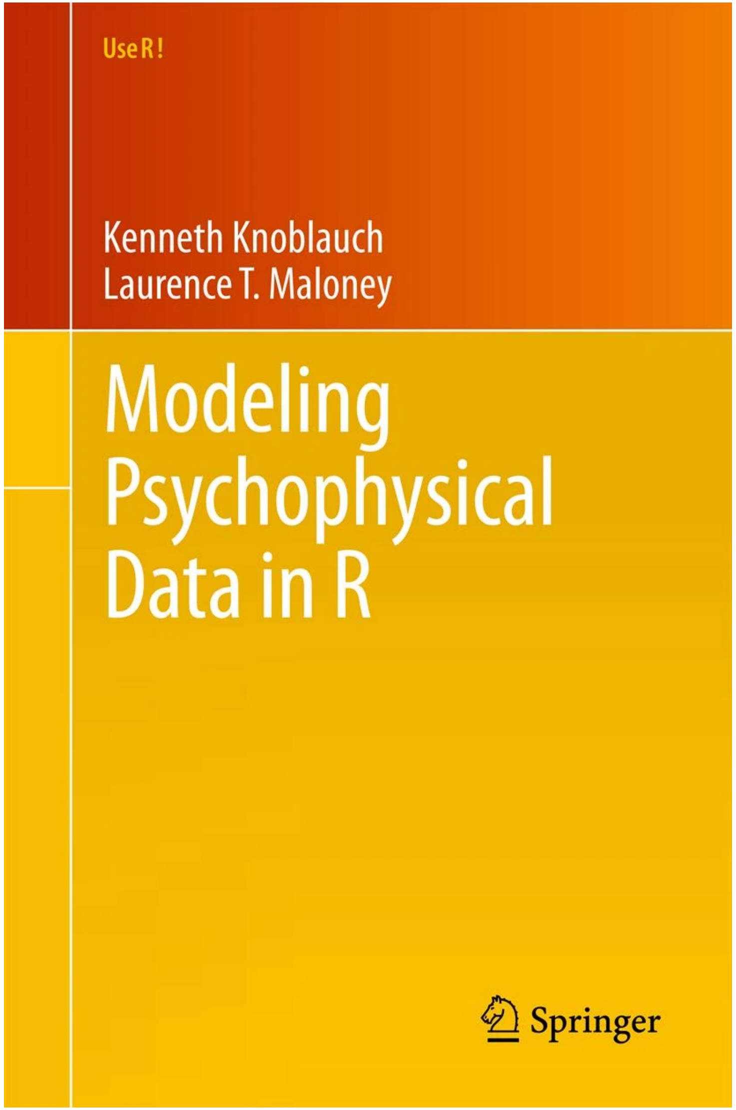
Methods to measure **supra-threshold appearance** of stimuli

Outcome: **perceptual scales**

	MLDS	**MLCM**
Use case	1 stimulus dimension	> 1 stimulus dimension
Task	Triads or quadruples	Paired comparisons
Observer model	Double differencing	Simple differencing
Estimation	Maximum likelihood via binomial GLM	

Thank you

References



Knoblauch & Maloney (2012)

References

Charrier, Maloney, Cherifi, & Knoblauch. (2007). Maximum likelihood difference scaling of image quality in compression-degraded images. JOSA A. doi: 10.1364/JOSAA.24.003418

Cheeseman et al. (2021). Scaling and discriminability of perceived gloss. JOSA A. doi:10.1364/josaa.409454

Gerardin, Devinck, Dojat, & Knoblauch. (2014). Contributions of contour frequency, amplitude, and luminance to the watercolor effect estimated by conjoint measurement. Journal of Vision. doi: 10.1167/14.4.9

Ho, Landy & Maloney (2008). Conjoint measurement of gloss and surface texture. Psychological Science. doi: 10.1111/j.1467-9280.2008.02067.x

Kingdom & Prins (2016). Psychophysics: A Practical Introduction. Academic Press.

Krantz et al. (1971). Foundations of Measurement Vol I. Dover Publications.

Maloney & Yang (2003). Maximum likelihood difference scaling. JoV. doi:10:1167/3.8.5

Shooner & Mullen (2022). Linking perceived to physical contrast: Comparing results from discrimination and difference-scaling experiments. JoV. doi:10.1167/JOV.22.1.13

Vincent, Maertens , & Aguilar. (2024). What Fechner could not do: Separating perceptual encoding and decoding with difference scaling. JoV. doi:10.1167/jov.24.5.5

White(1979). A new effect of pattern on perceived lightness. Perception. doi:10.1068/p080413

Wiebel, Aguilar, & Maertens. (2017). Maximum likelihood difference scales represent perceptual magnitudes and predict appearance matches. JoV. doi:10.1167/17.4.1