Models of metacognition

Computational Psychiatry Course – Zurich 2022

Dr. Marion ROUAULT (marion.rouault@gmail.com)

- Ecole Normale Supérieure de Paris
- Paris Brain Institute

With thanks to O. Harrison and S. Fleming

What is metacognition and why is it useful?

- “cognition about cognitive phenomena...” Flavell, 1979

- “the processes by which people self-reflect on their own cognitive and memory processes (**monitoring**) and how they put their metaknowledge to use in regulating their information processing and behaviour (**control**)” Koriat, 2007

Daily examples of metacognition

Will I be able to learn this topic?

How confident am I in my decision?

I can't remember it now, but I know it when I see it

I'm driving too fast, I feel out of control

Did I really speak to my partner last night or was I dreaming?

A definition

- Most perceptions, memories and choices are accompanied by subjective estimates of their reliability i.e. **confidence** estimates

  Kepecs et al., 2008; Lebreton et al., 2015; Arango-Muños & Bermúdes, 2018

A definition

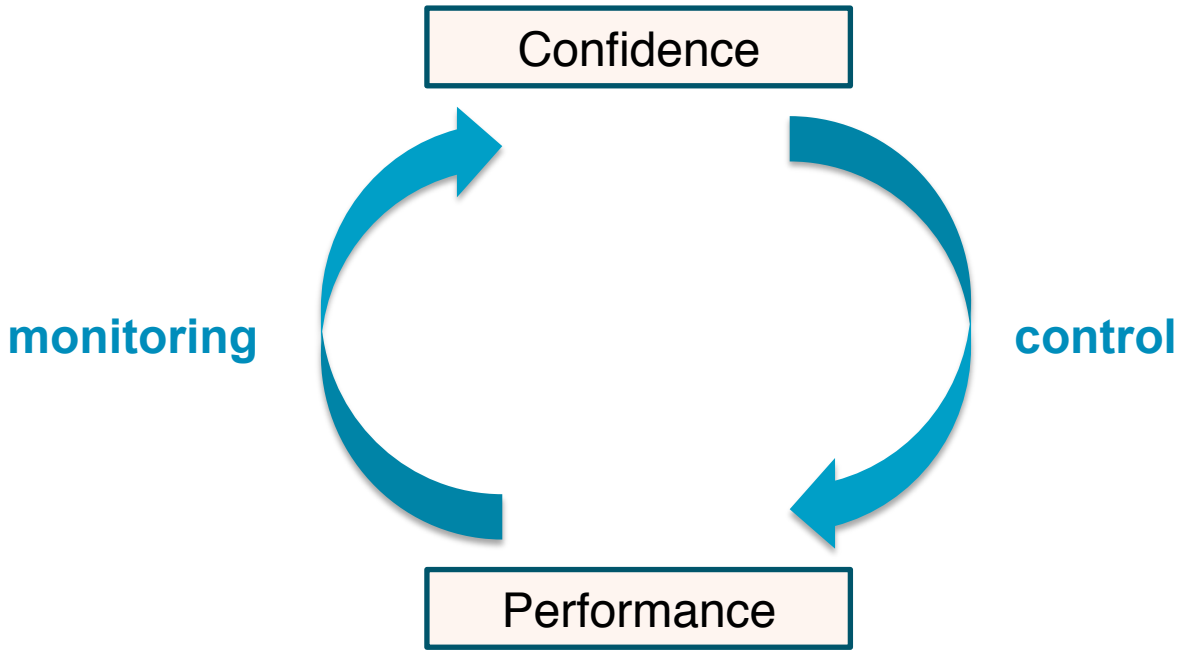
- Most perceptions, memories and choices are accompanied by subjective estimates of their reliability i.e. **confidence** estimates

  Kepecs et al., 2008; Lebreton et al., 2015; Arango-Muños & Bermúdes, 2018

- **Metacognition** refers to our ability to **monitor**, **evaluate**, and **reflect on** our own cognitive processes

  Fleming & Dolan, 2012; Gehring et al., 1993

Reciprocal interactions between cognition and metacognition



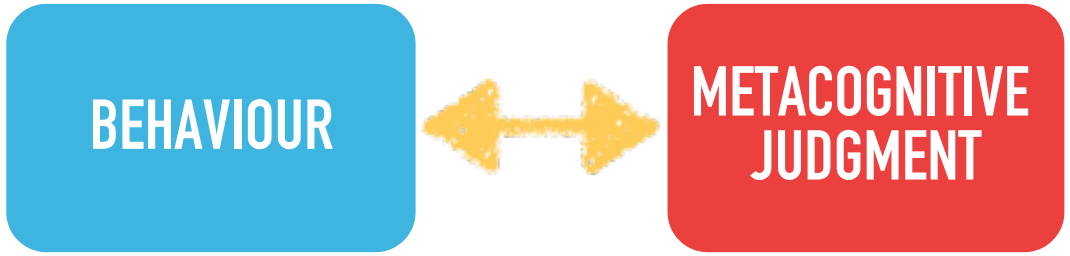
Why study metacognition?

Inaccurate metacognitive knowledge of cognitive and physical impairments is common in **psychiatric** and **neurological** disorders and in **healthy aging**

Insight = the capacity for accurate metacognition

Metacognition: a central aspect of neurological and psychiatric pathologies

A primer on measuring metacognition

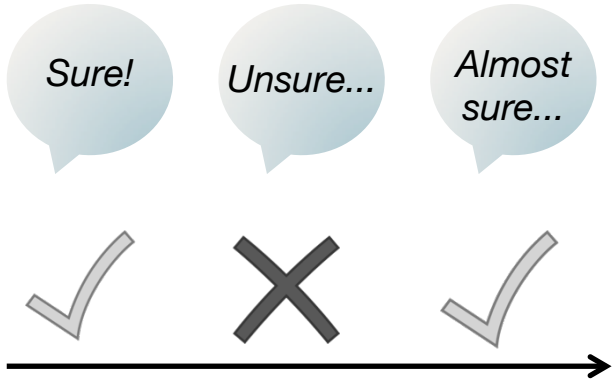


E.g. answer to exam question; response in a psychophysics experiment

E.g. **confidence** in getting the answer right

Metacognition in the lab

- Not possible to assess metacognition from a single judgment
- Need multiple judgments over time, examine **statistical association** between behavioural responses and metacognitive judgments



	High confidence	Low confidence
Correct		
Incorrect		

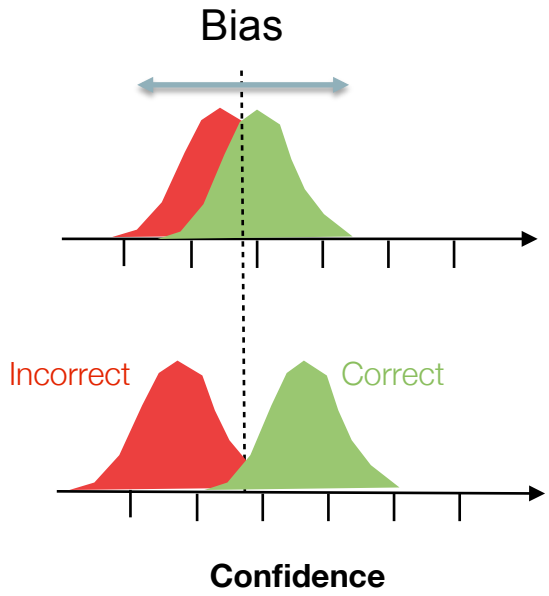
High metacognitive ability
Low metacognitive ability

Metacognitive bias and sensitivity

Confidence in behaviour

Metacognitive bias

Metacognitive sensitivity



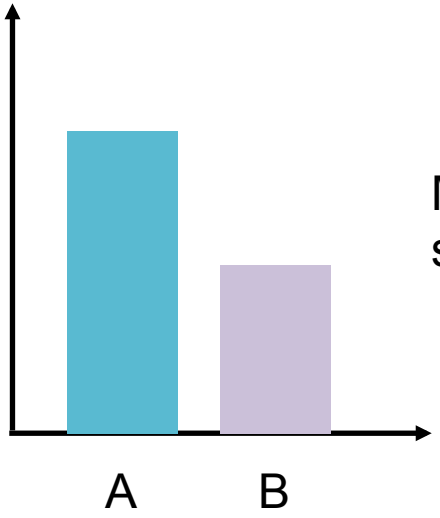
Other terminology in the literature:

Bias: calibration, confidence level, self-perceived ability, self-belief
Sensitivity: discrimination, resolution, metacognitive awareness, insight

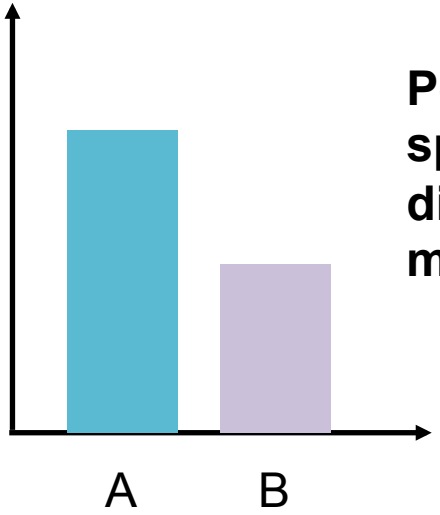
Fleming & Lau, 2014 *Frontiers*

Importance of taking into account performance

First-order performance

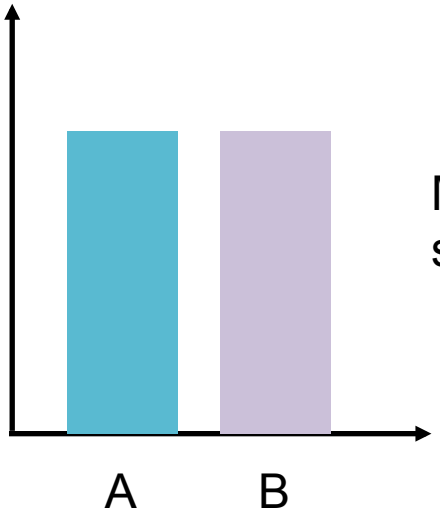


Metacognitive sensitivity

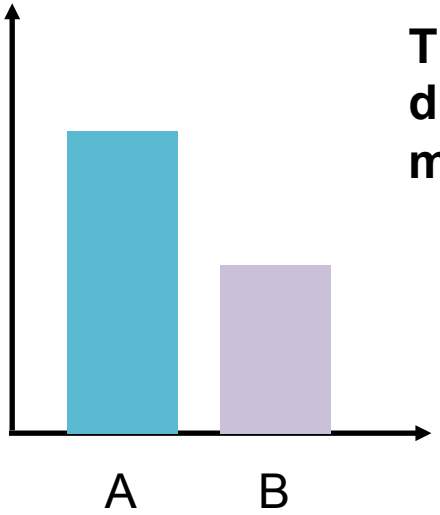


Possibly spurious difference in metacognition

First-order performance



Metacognitive sensitivity



True difference in metacognition

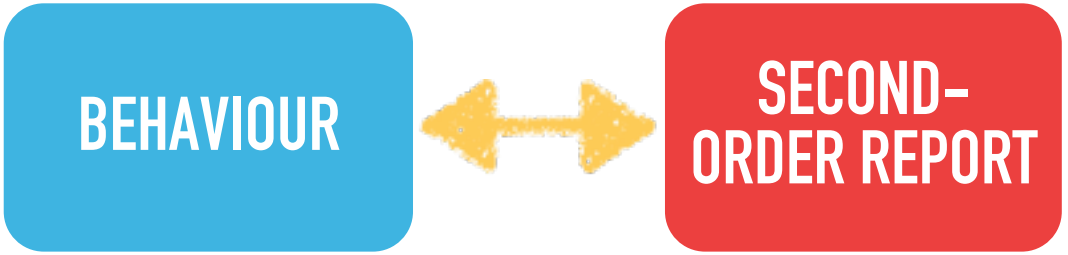
How to measure metacognitive sensitivity?

Ideal measure: should identify differences in metacognitive sensitivity, but be unaffected by metacognitive bias (overall confidence) or task performance

3 main approaches:

1. Correlation approaches
2. Area under type 2 ROC (AUROC2)
3. Meta-*d'*

Quantifying metacognition (1) - correlation approaches



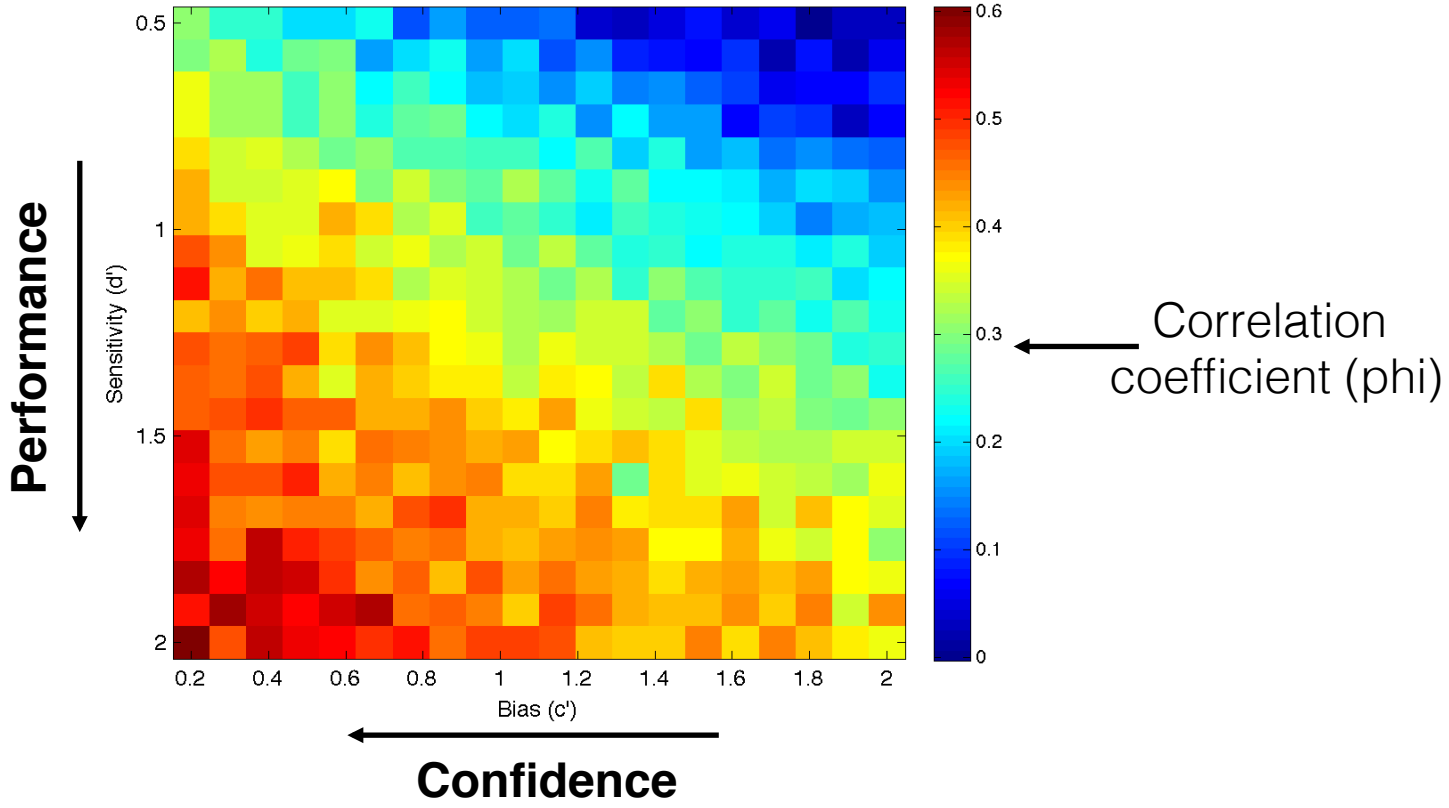
	High confidence	Low confidence
Correct	A	B
Incorrect	C	D

Decision = [1 0 0 1 1 0 1 0…]

Phi = corr(decision, confidence)

Confidence = [0 0 0 1 1 1 1 0…]

Problems with correlation approaches



Simple measures of association (gamma, phi) are not bias free, and confounded by performance (see Mason & Rotello, 2009; Fleming & Lau, 2014)

Should I use correlation measures?

- **Pros:**
  - Very simple, can be used with pretty much any design (just need to define trials as correct vs. incorrect)
  - Useful if one wants to establish presence (vs. absence) of metacognition

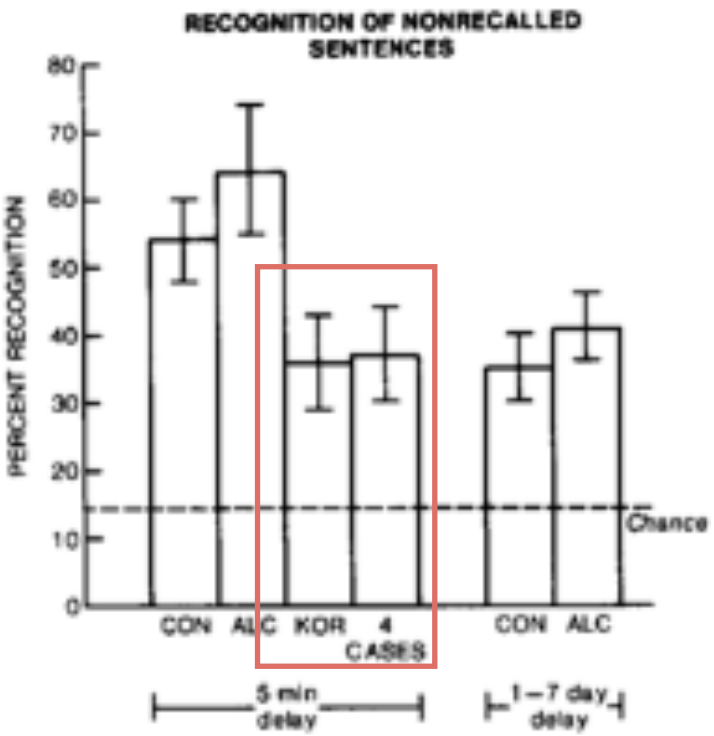
- **Cons:**
  - Confounded by performance
  - Confounded by confidence bias
  - Need to be careful in matching these influences if comparing phi/gamma between conditions

Journal of Experimental Psychology:
Learning, Memory, and Cognition
1986, Vol. 12, No. 3, 452-460

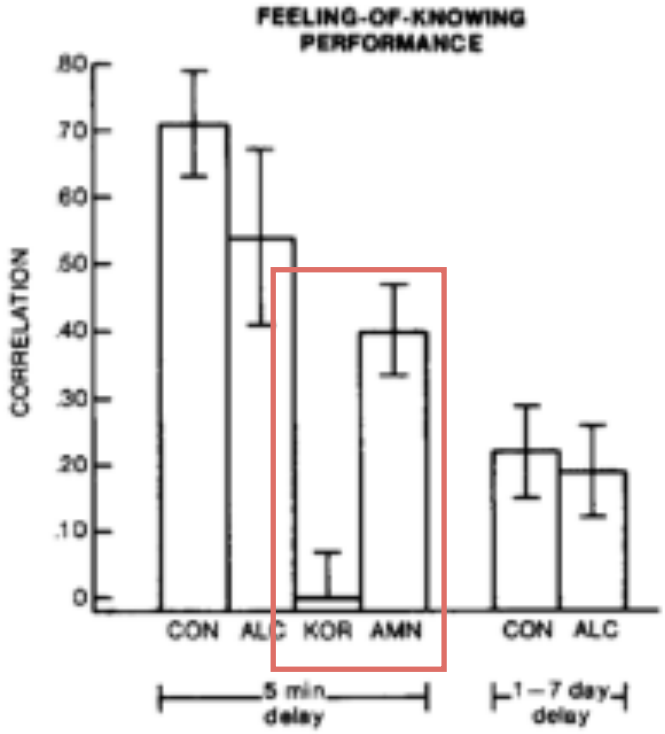
0278-7393/86/$00.75

Memory and Metamemory: A Study of the Feeling-of-Knowing Phenomenon in Amnesic Patients

Arthur P. Shimamura and Larry R. Squire
Veterans Administration Medical Center, San Diego and Department of Psychiatry, University of California, San Diego, School of Medicine



Task performance



Metacognitive sensitivity (performance-confidence correlation)

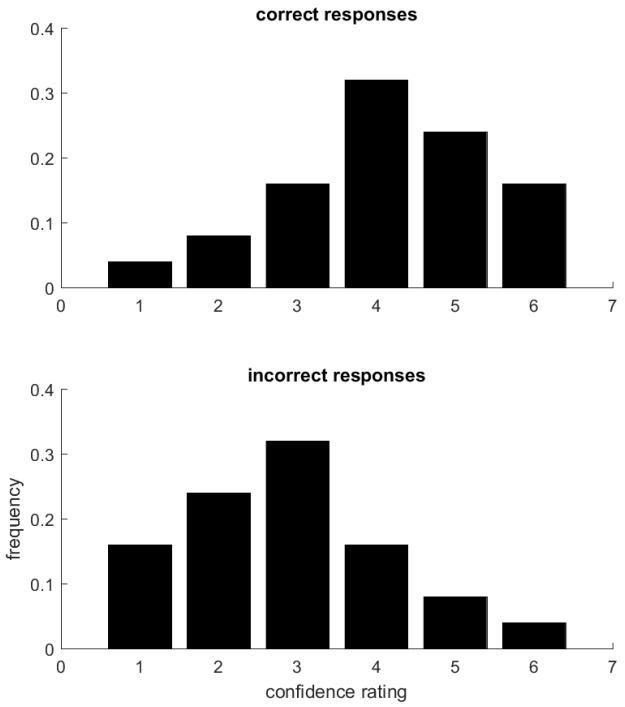
Quantifying metacognition (2) - AUROC2

Two Types of ROC Curves and Definitions of Parameters*

F. R. Clarke, T. G. Birdsall, and W. P. Tanner, Jr.
Electronic Defense Group, University of Michigan, Ann Arbor, Michigan
(Received February 26, 1959)

Type 2 receiver operating characteristic (ROC) curves are a compact representation of metacognitive sensitivity

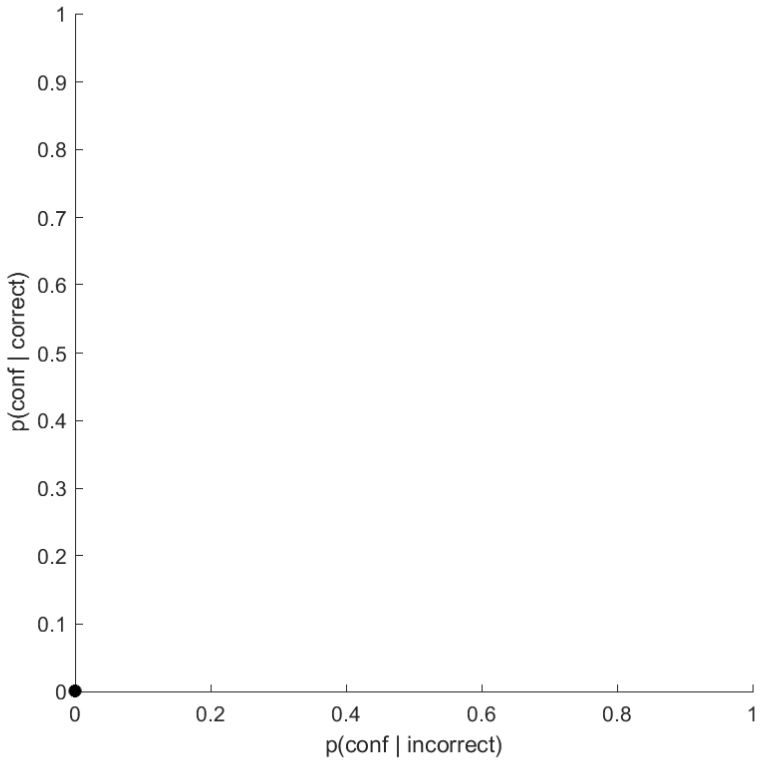
In general, the more distinct the confidence distributions for correct and for incorrect responses are, the more insight one has into the quality of individual decisions

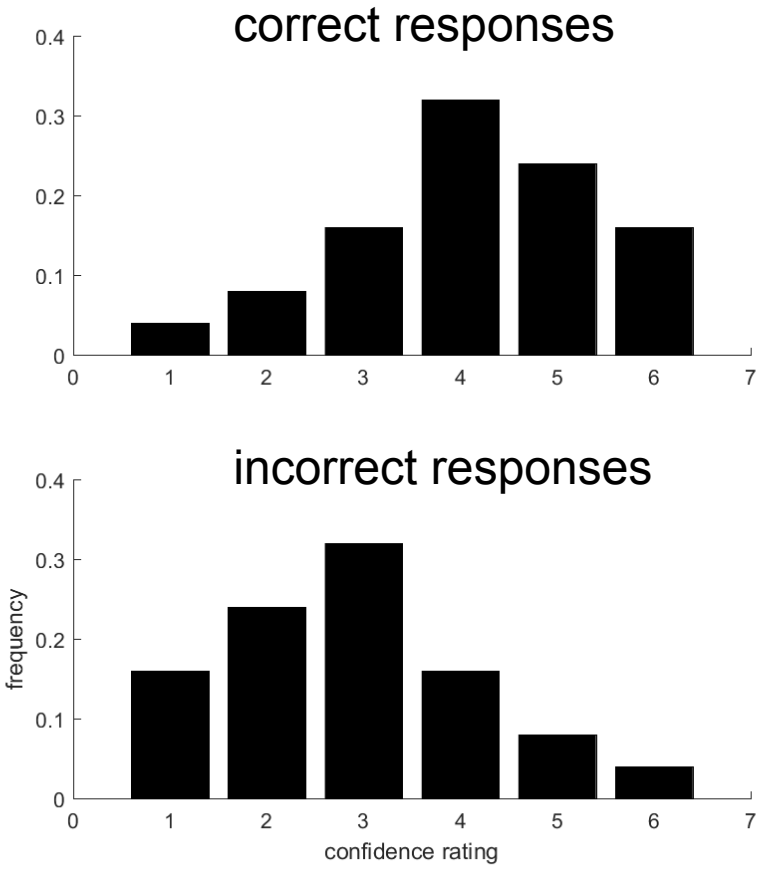


with thanks to Matan Mazor

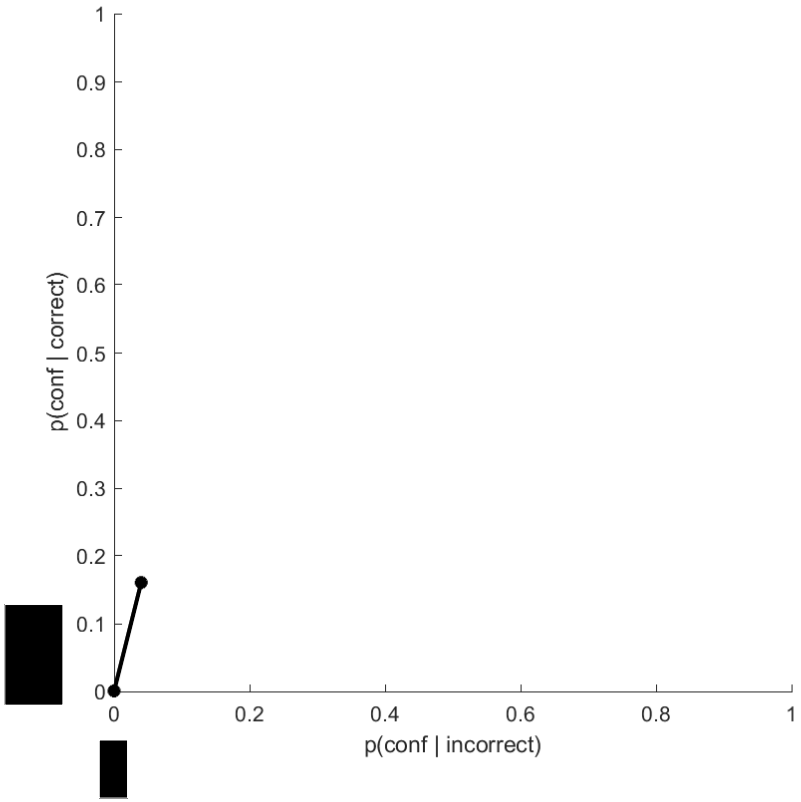
Clarke, Birdsall & Tanner (1959) *J Acoust Soc Am*; Galvin et al. (2003) *Psych Bull Rev*

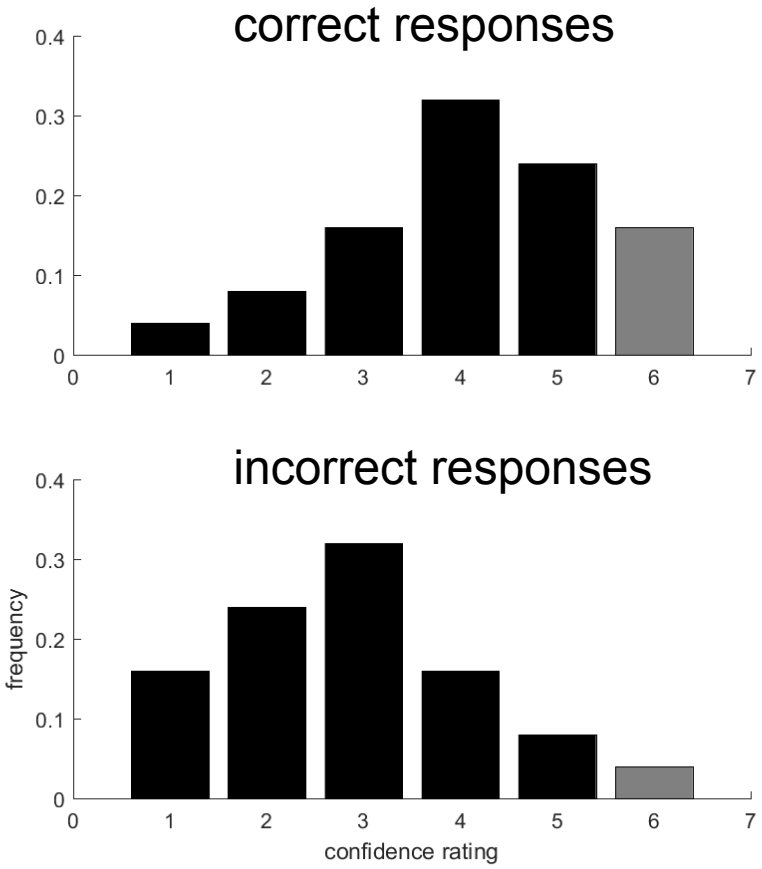
Type 2 ROCs


p(conf | correct)
p(conf | incorrect)
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
0.2
0.4
0.6
0.8
1

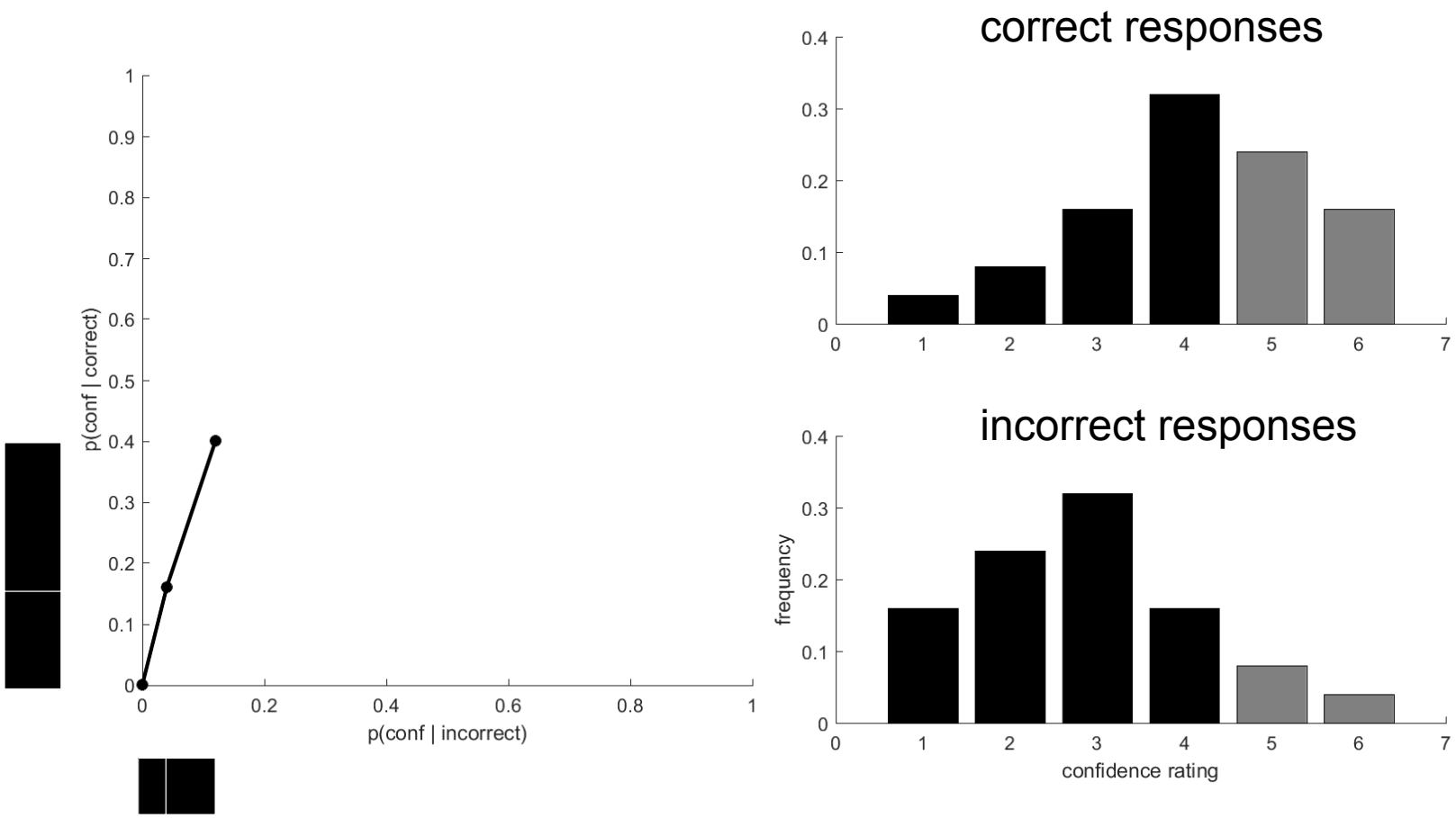

correct responses
incorrect responses
frequency
confidence rating
0
0.1
0.2
0.3
0.4
0
1
2
3
4
5
6
7

Type 2 ROCs

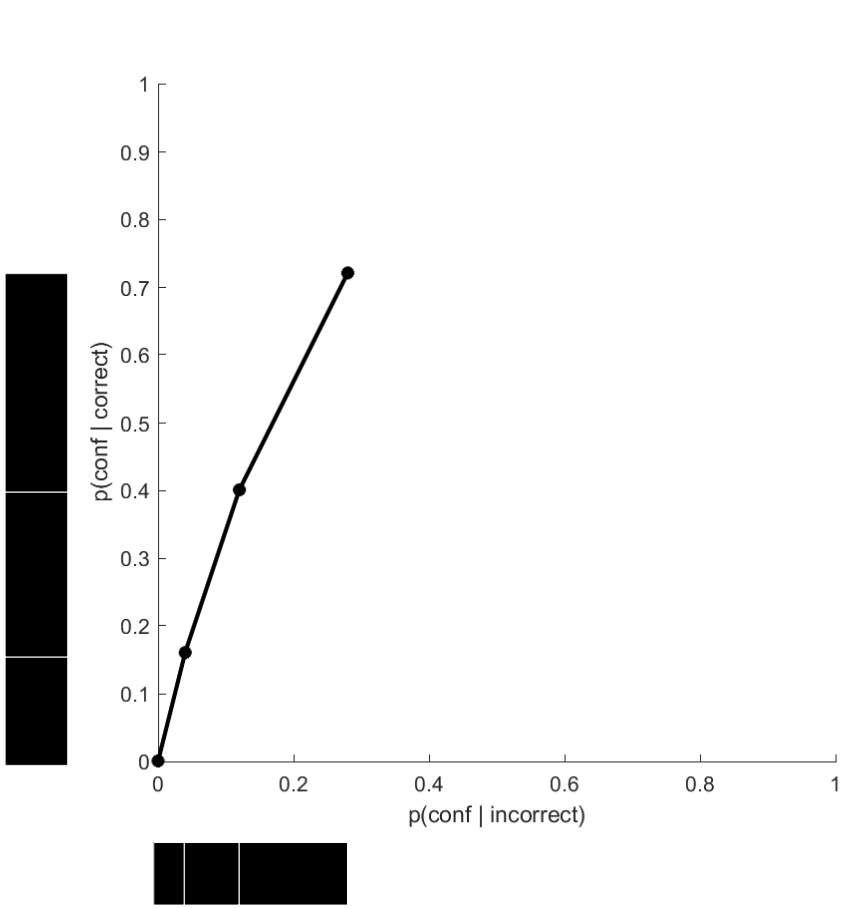

p(conf | correct)
p(conf | incorrect)
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
0.2
0.4
0.6
0.8
1

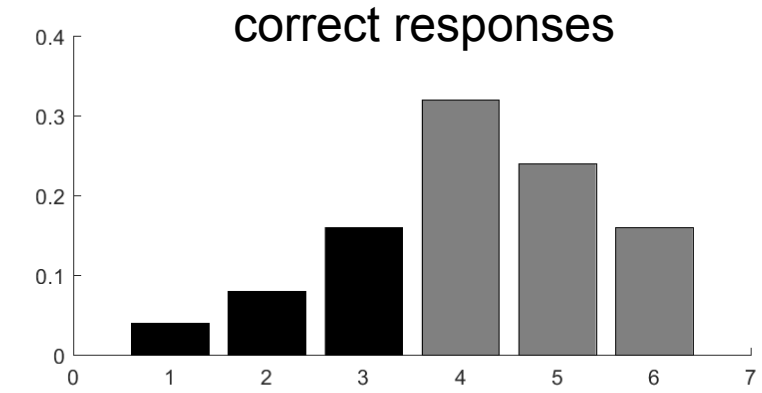

correct responses
incorrect responses
frequency
confidence rating
0
0.1
0.2
0.3
0.4
0
1
2
3
4
5
6
7

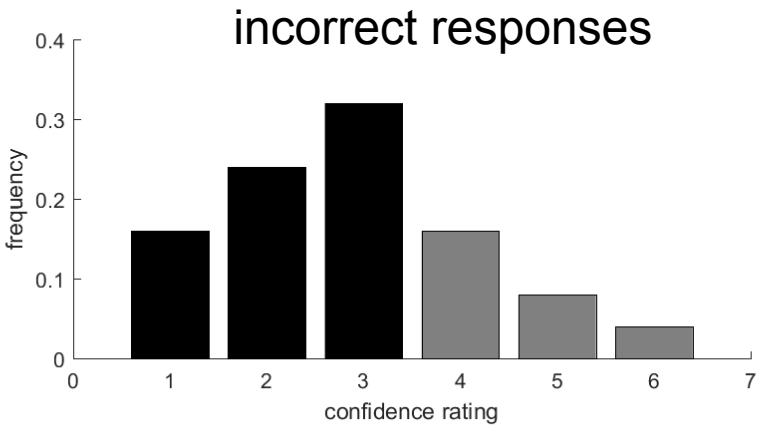
Type 2 ROCs


p(conf | correct)
p(conf | incorrect)
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0.2
0.4
0.6
0.8
correct responses
incorrect responses
frequency
confidence rating
0
1
2
3
4
5
6
7

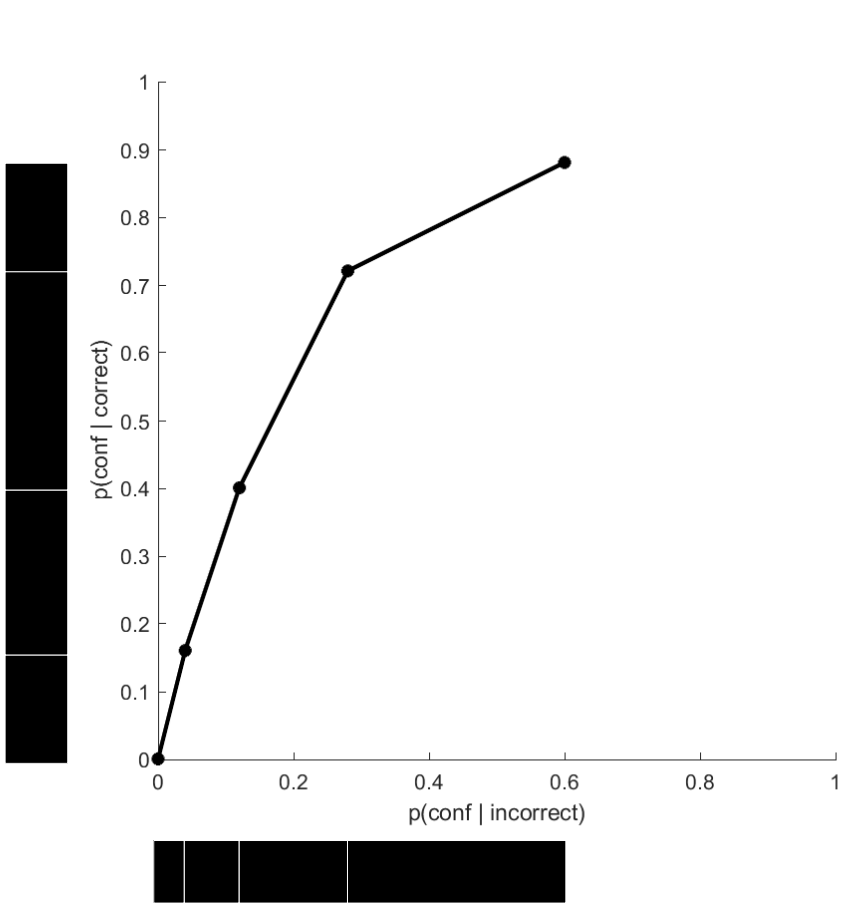
Type 2 ROCs

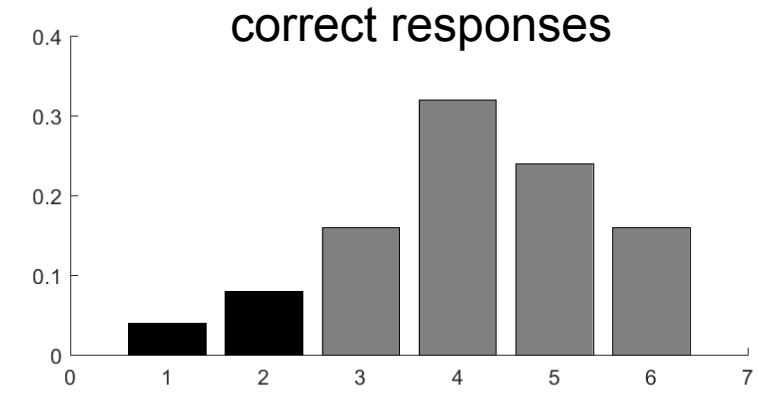

p(conf | correct)
p(conf | incorrect)
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0.2
0.4
0.6
0.8

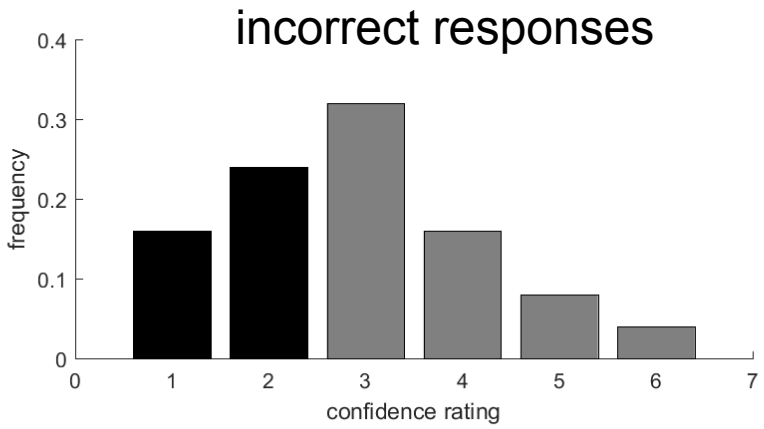

correct responses
0.1
0.2
0.3
0.4
0
1
2
3
4
5
6
7


incorrect responses
frequency
confidence rating
0.1
0.2
0.3
0.4
0
1
2
3
4
5
6
7

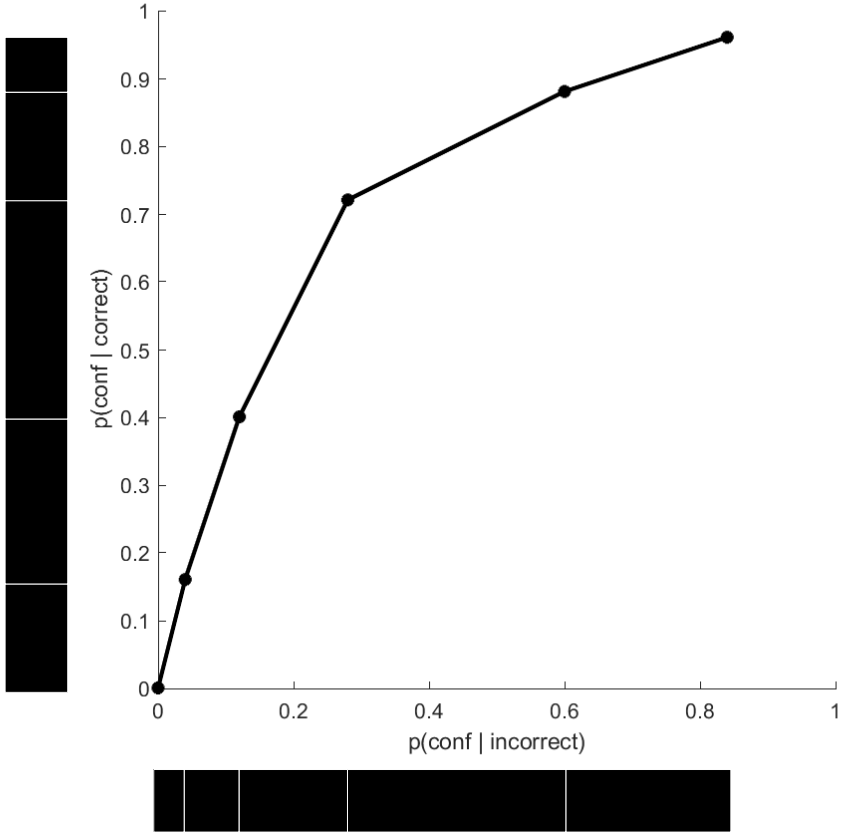
Type 2 ROCs

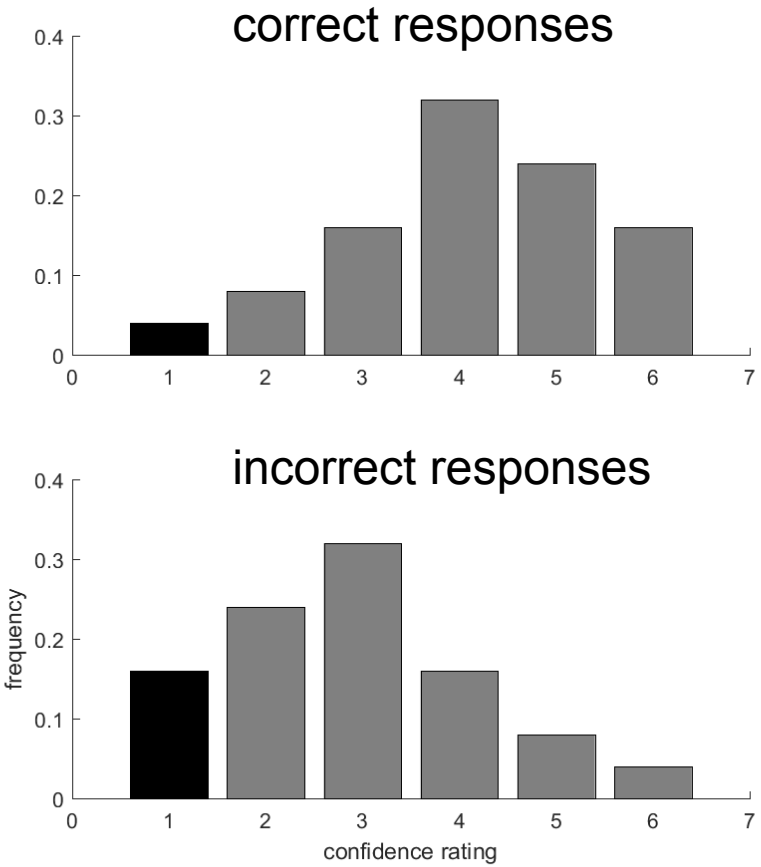

p(conf | correct)
p(conf | incorrect)
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
0.2
0.4
0.6
0.8
1


correct responses
0.4
0.3
0.2
0.1
0
0
1
2
3
4
5
6
7

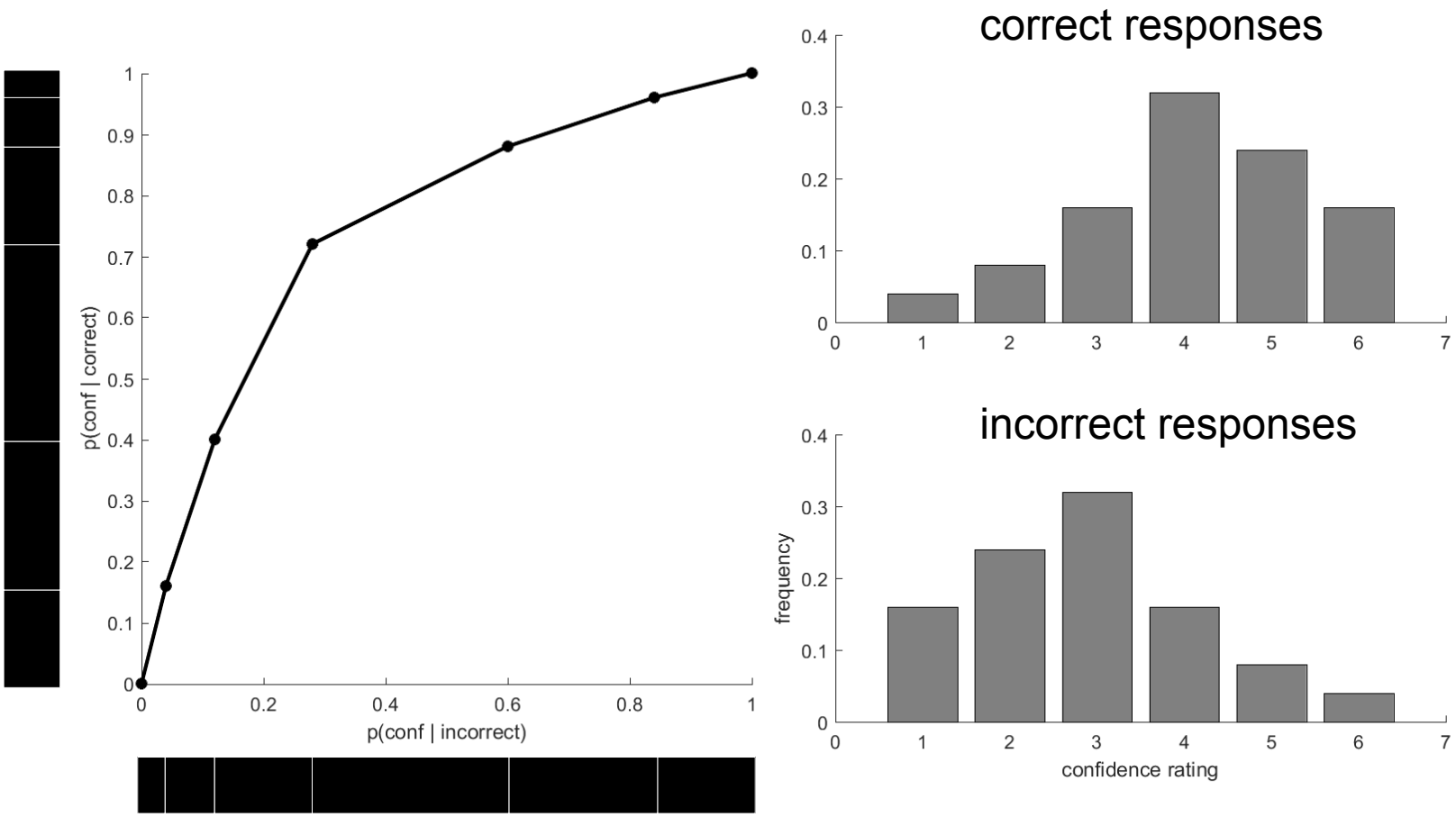

incorrect responses
frequency
0.4
0.3
0.2
0.1
0
0
1
2
3
4
5
6
7
confidence rating

Type 2 ROCs

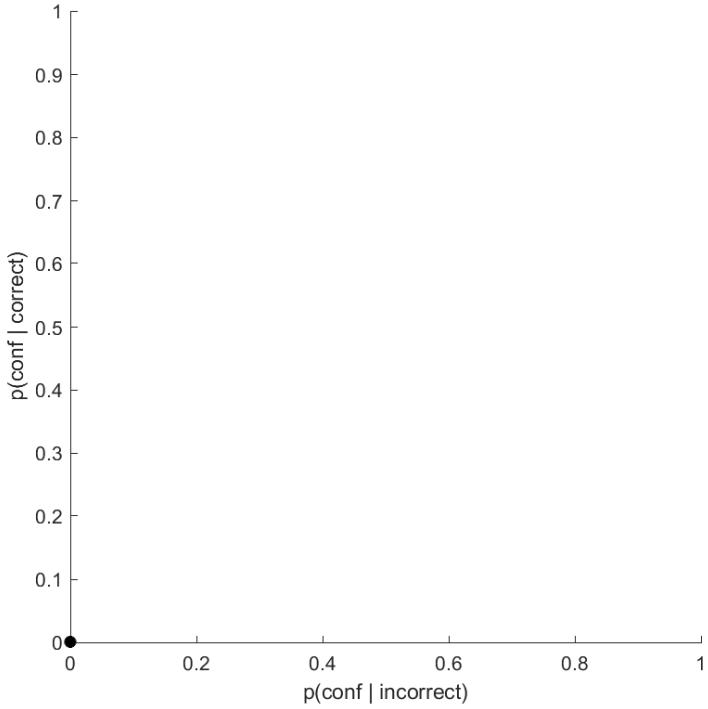

p(conf | correct)
p(conf | incorrect)
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
0.2
0.4
0.6
0.8
1


correct responses
incorrect responses
frequency
confidence rating
0
0.1
0.2
0.3
0.4
0
1
2
3
4
5
6
7

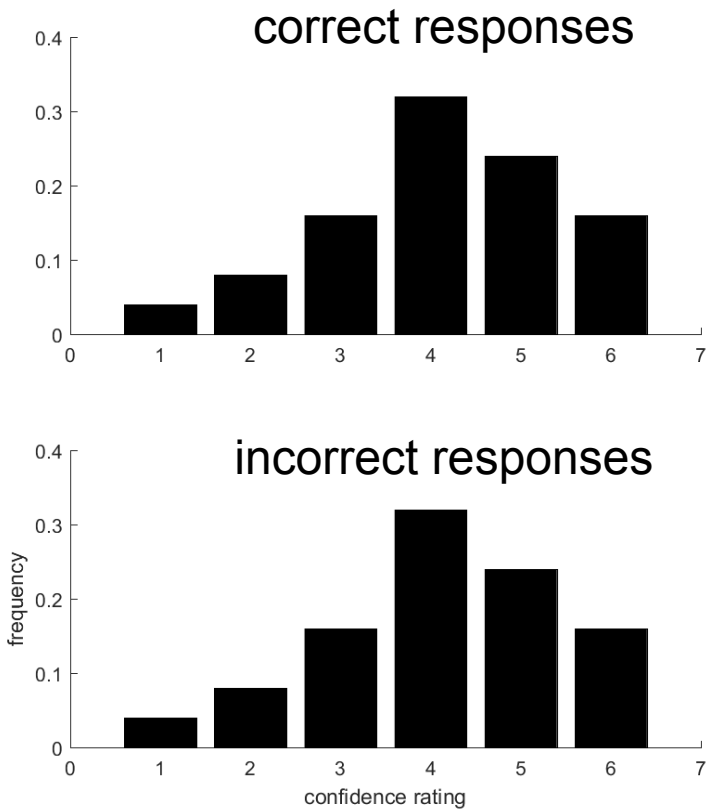
Type 2 ROCs


correct responses
incorrect responses
p(conf | correct)
p(conf | incorrect)
frequency
confidence rating
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
2
3
4
5
6
7

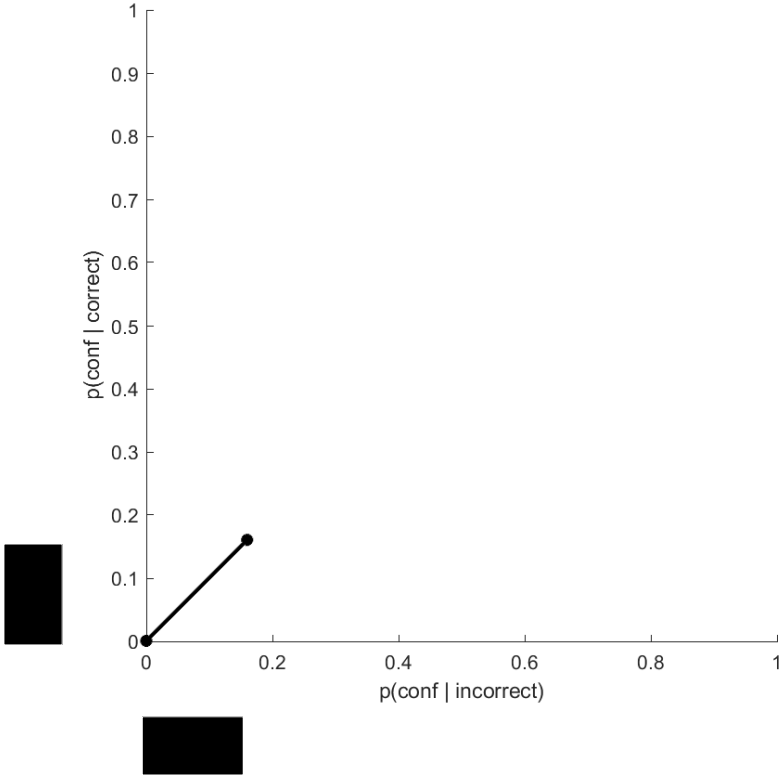
Type 2 ROCs



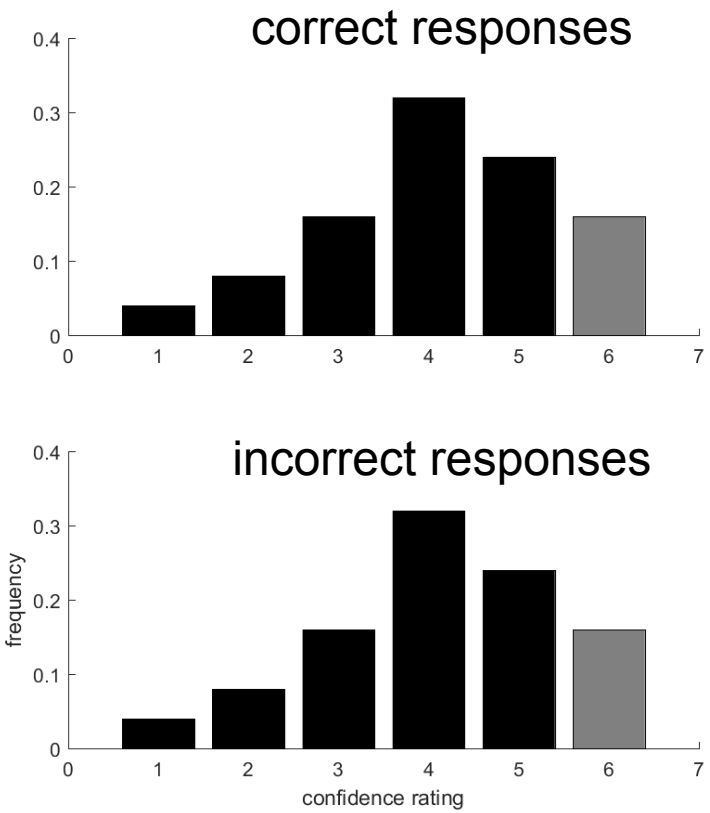
No metacognition:



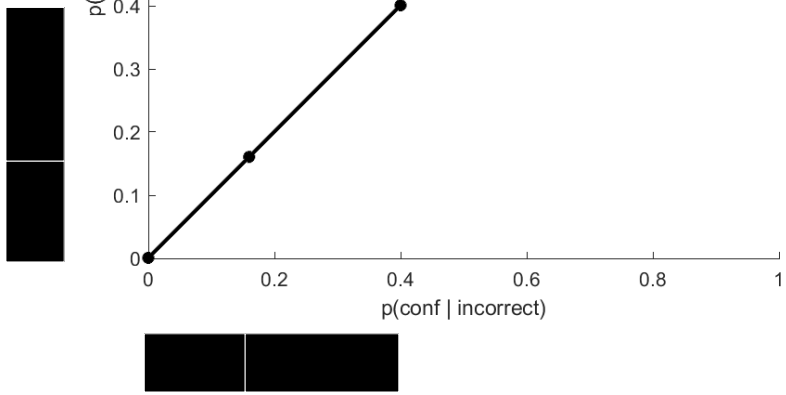
Type 2 ROCs



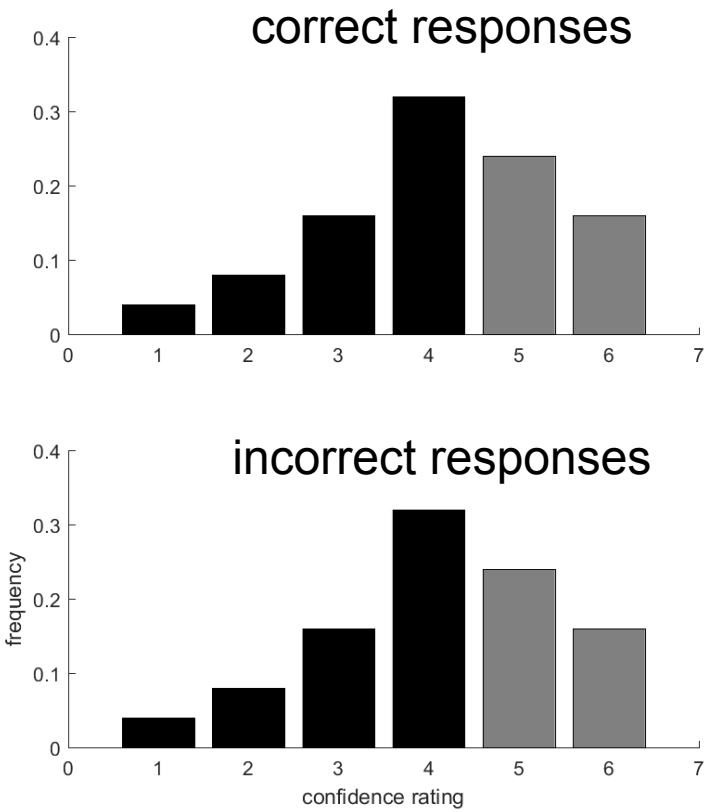
No metacognition:



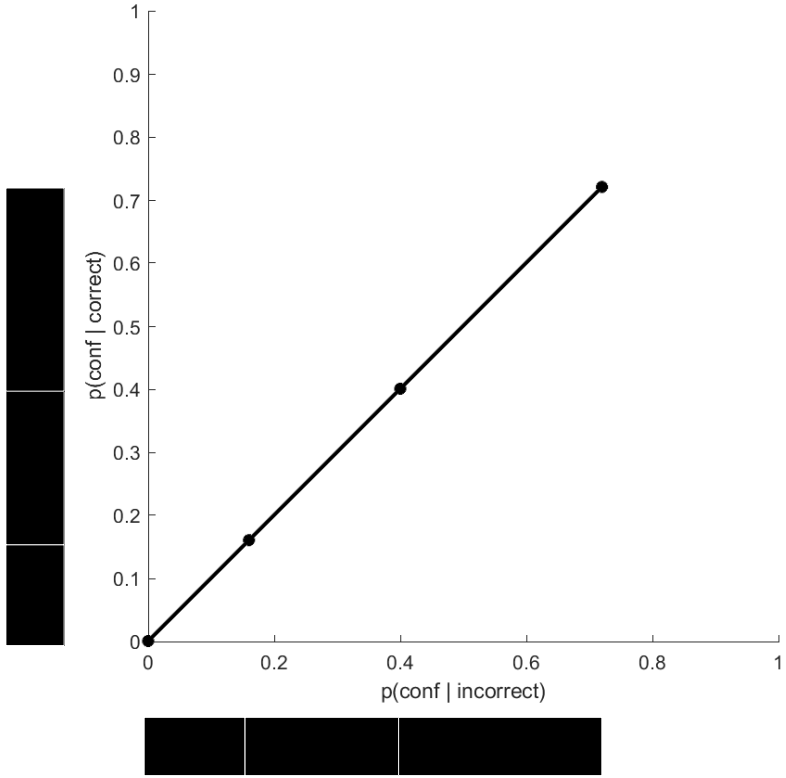
Type 2 ROCs



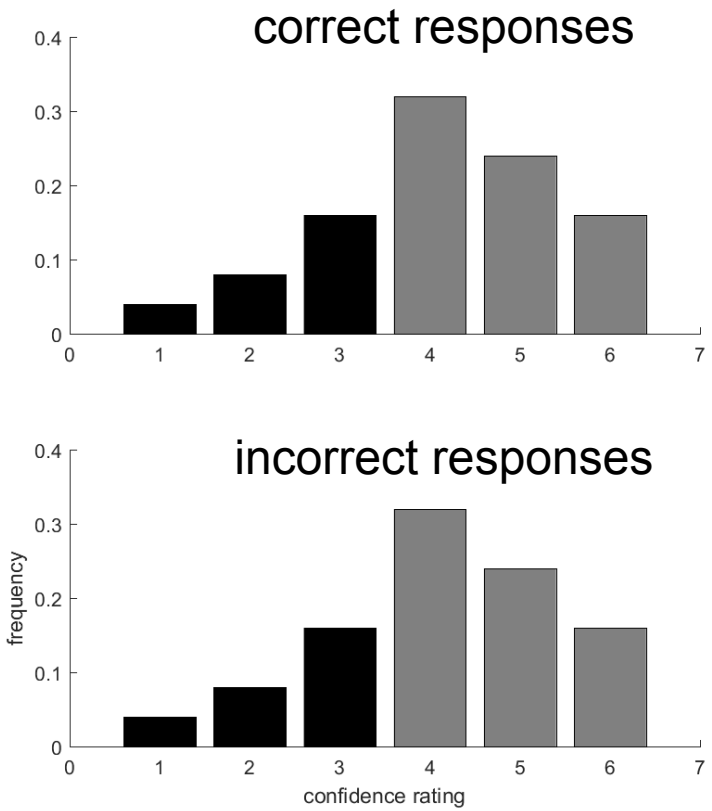
No metacognition:



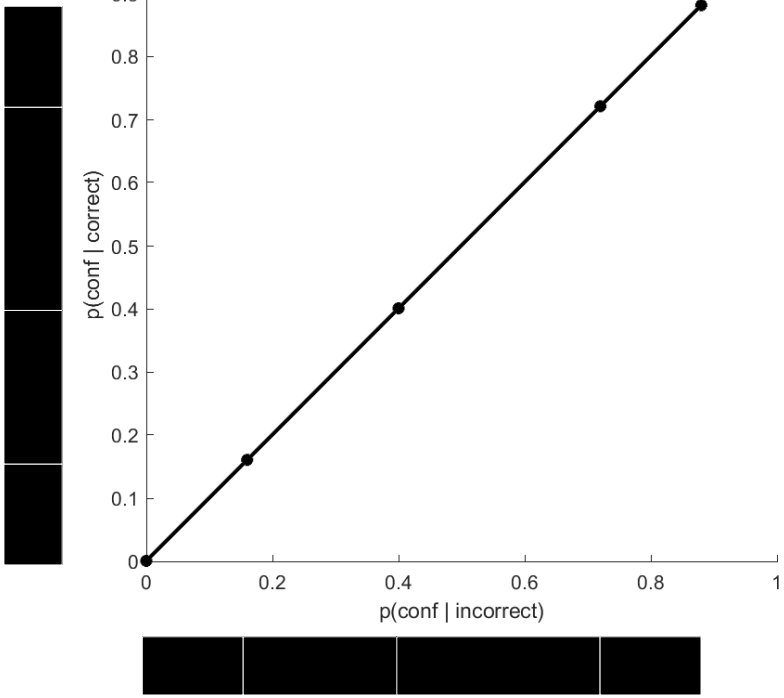
Type 2 ROCs



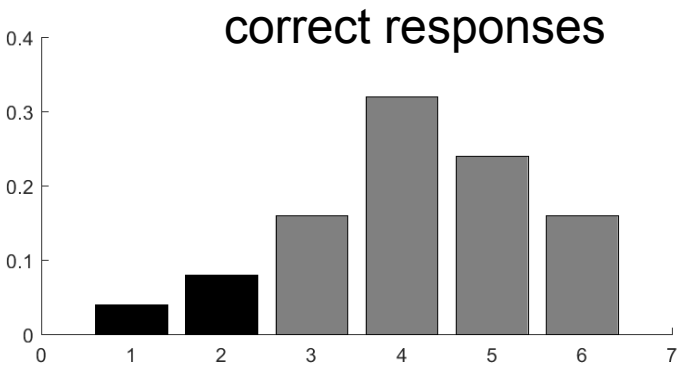
No metacognition:

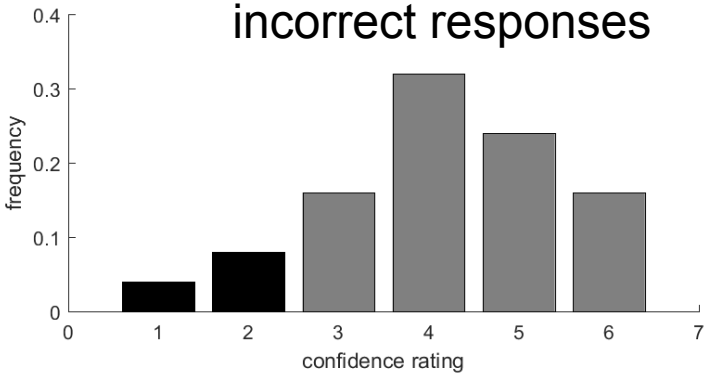


Type 2 ROCs

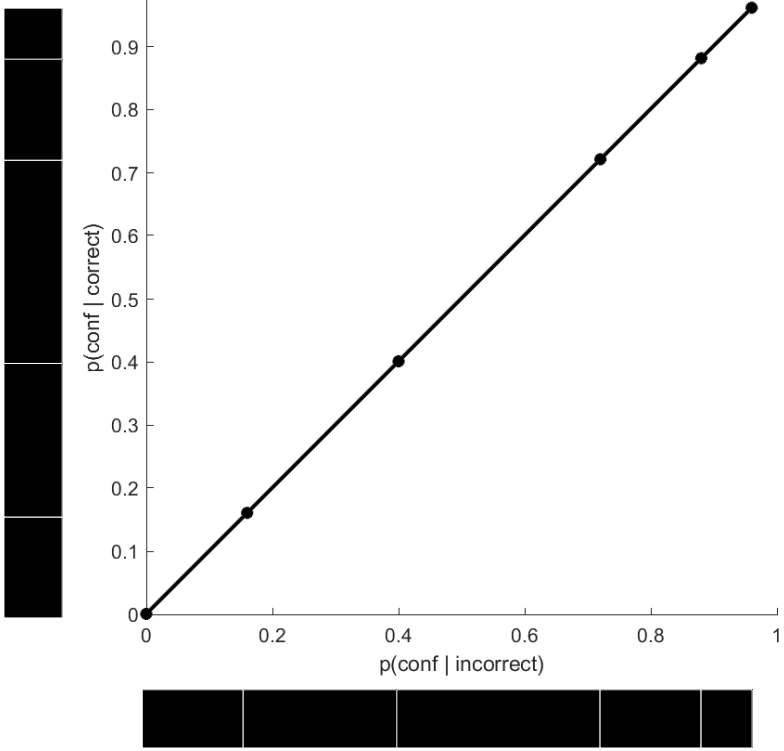


No metacognition:

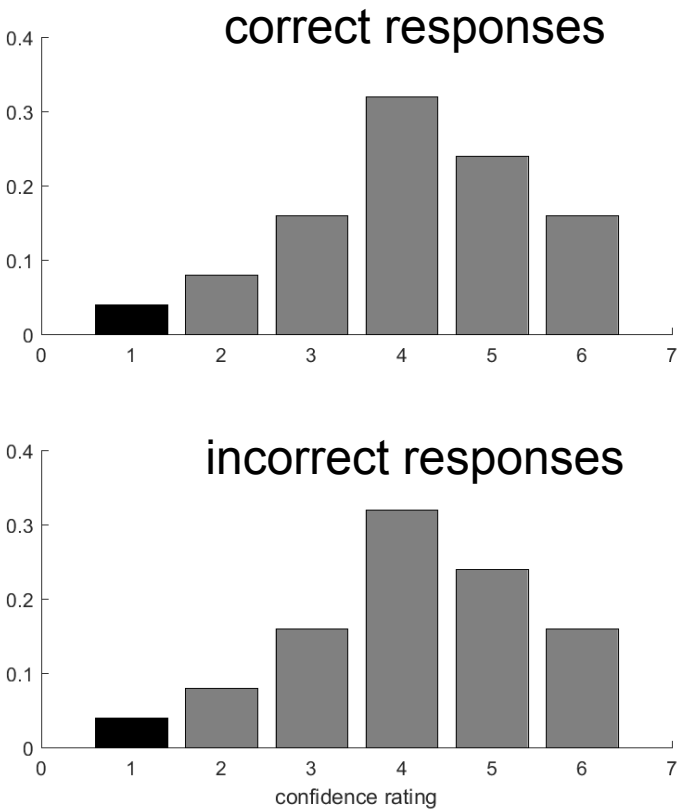




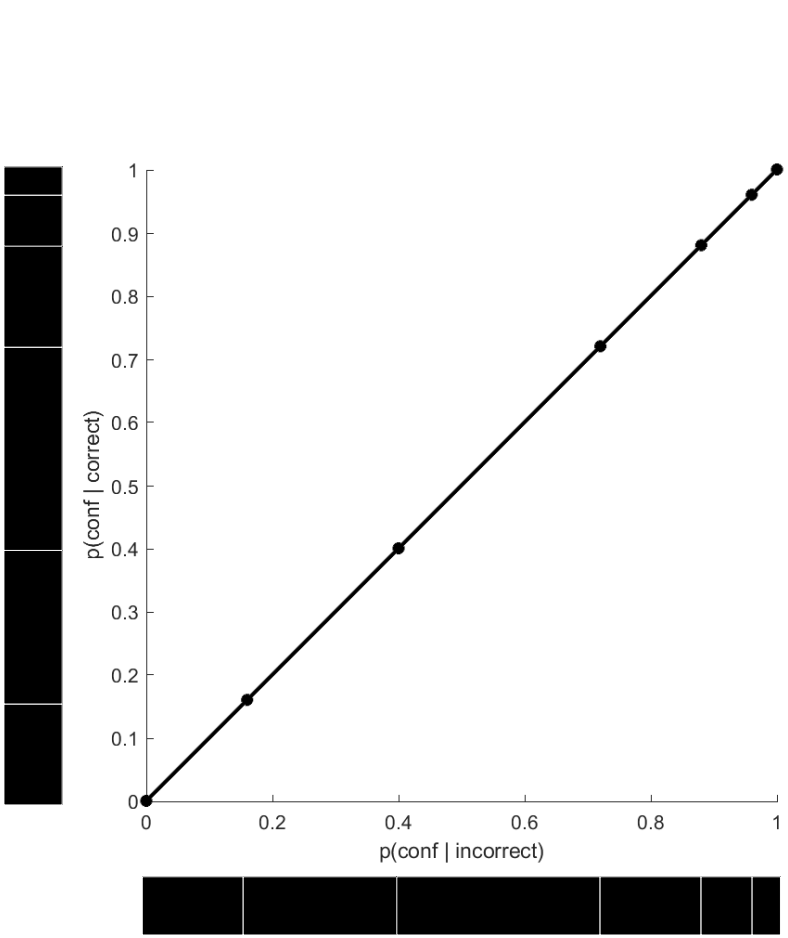
Type 2 ROCs



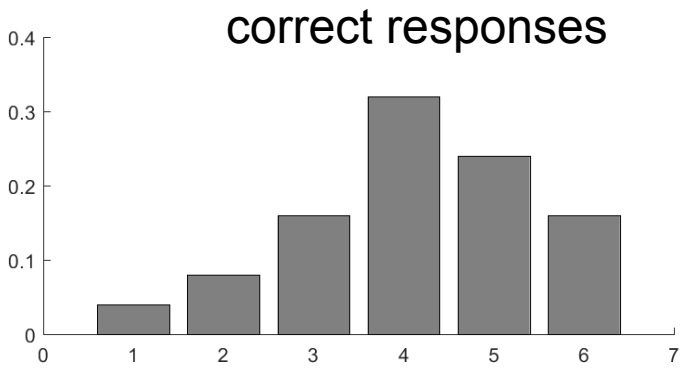
No metacognition:

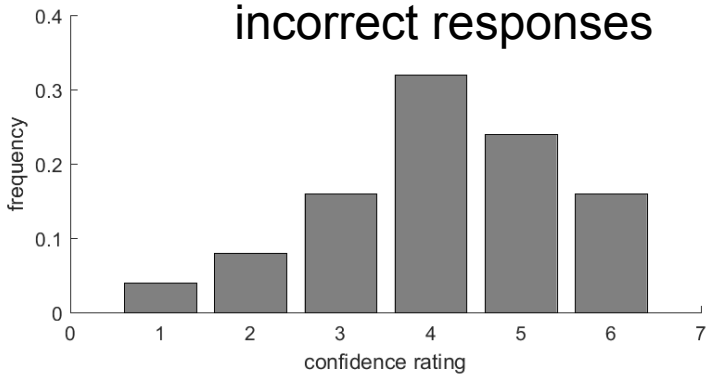


Type 2 ROCs



No metacognition:





Type 2 ROCs

- Simple measure of metacognitive sensitivity
- Theoretically independent of metacognitive bias (overall confidence)
- **BUT *not* independent of performance…**

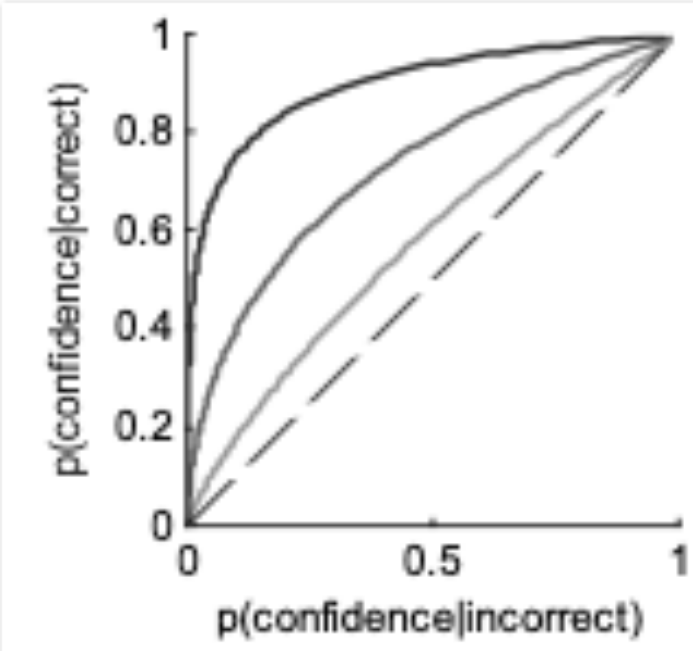
Simulated type 2 ROC

Type 1 performance

d' = 0.5

d' = 1.5

d' = 3.0

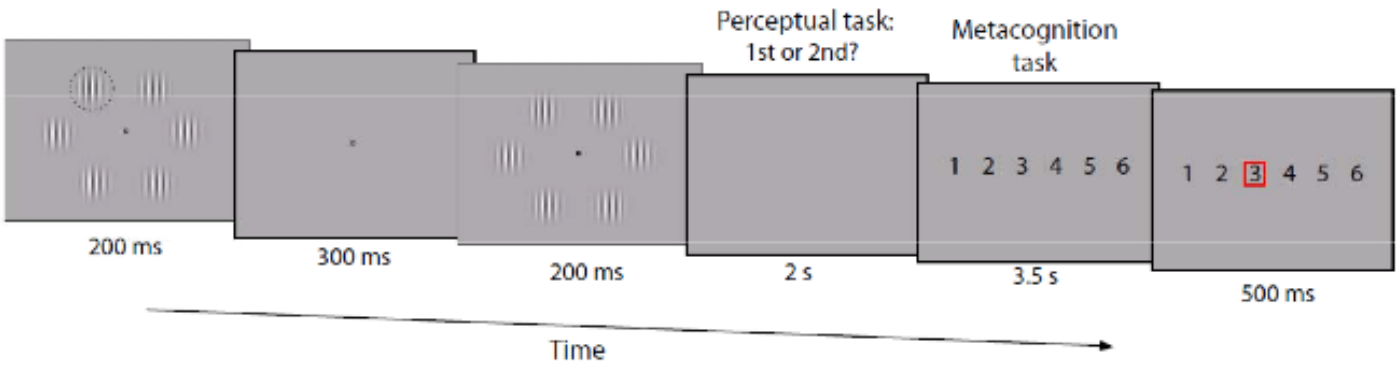


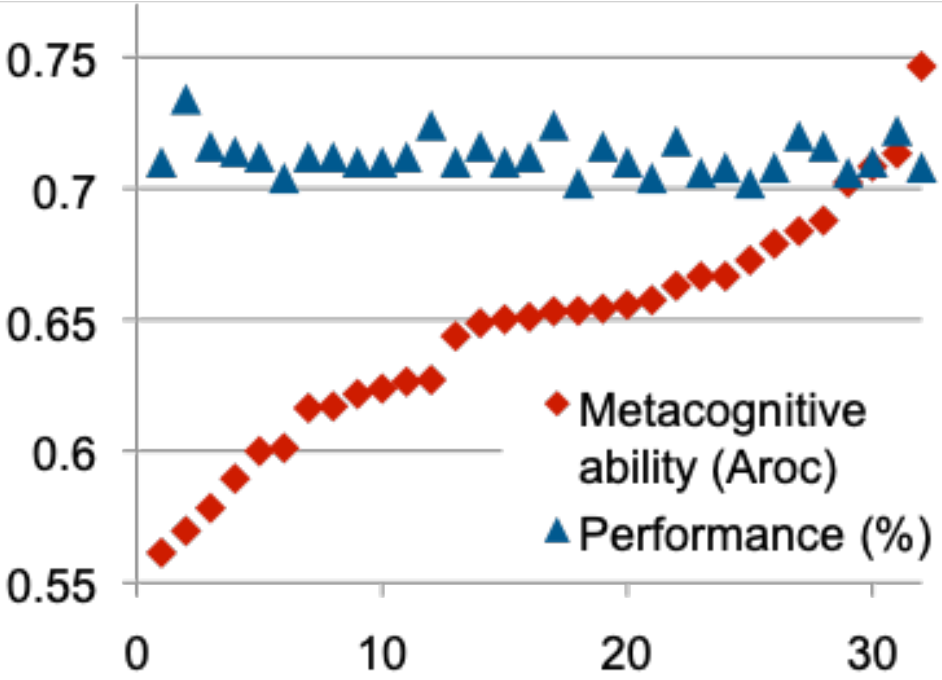
Should I use AUROC2?

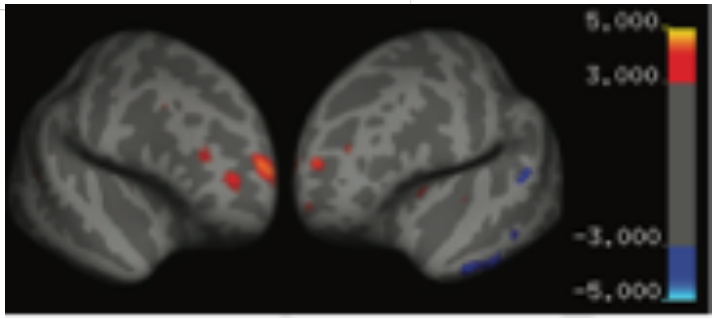
- **Pros:**
  - Non-parametric measure of sensitivity, can be used with pretty much any design (correct vs. incorrect)
  - Independent of metacognitive bias in most circumstances

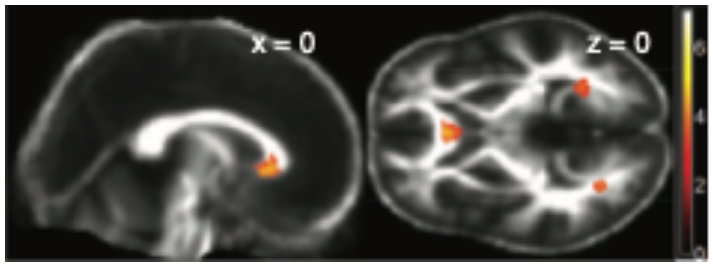
- **Cons:**
  - Confounded by performance
  - Also affected by response criterion (though not usually a major issue in many experimental designs
  - Needs performance-controlled paradigm, or control for performance in analysis (e.g. multiple regression)

Empirical example







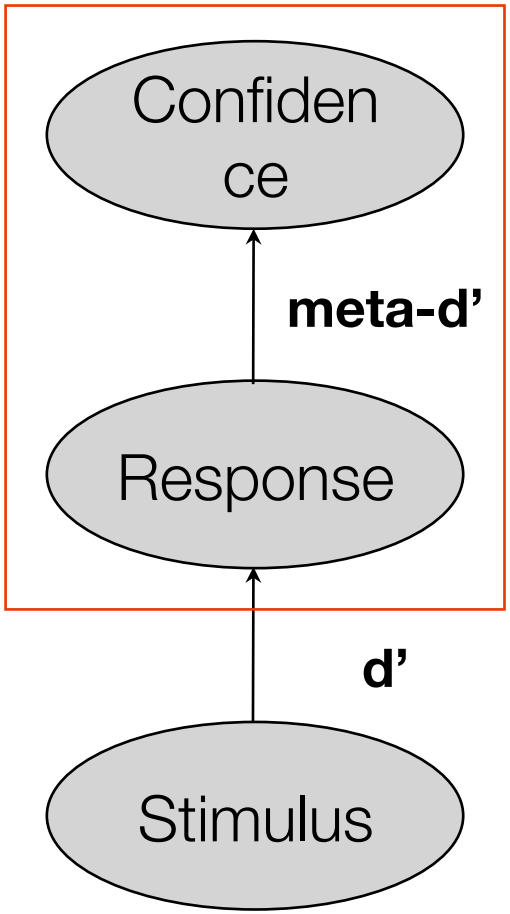


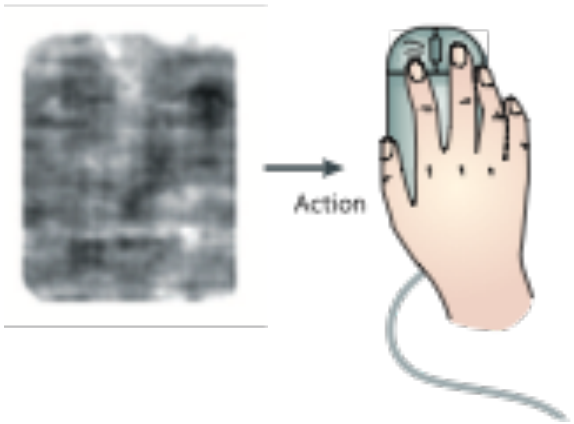
Fleming et al. (2010) *Science*

Quantifying metacognition (3) - meta-d’

Metacognitive sensitivity

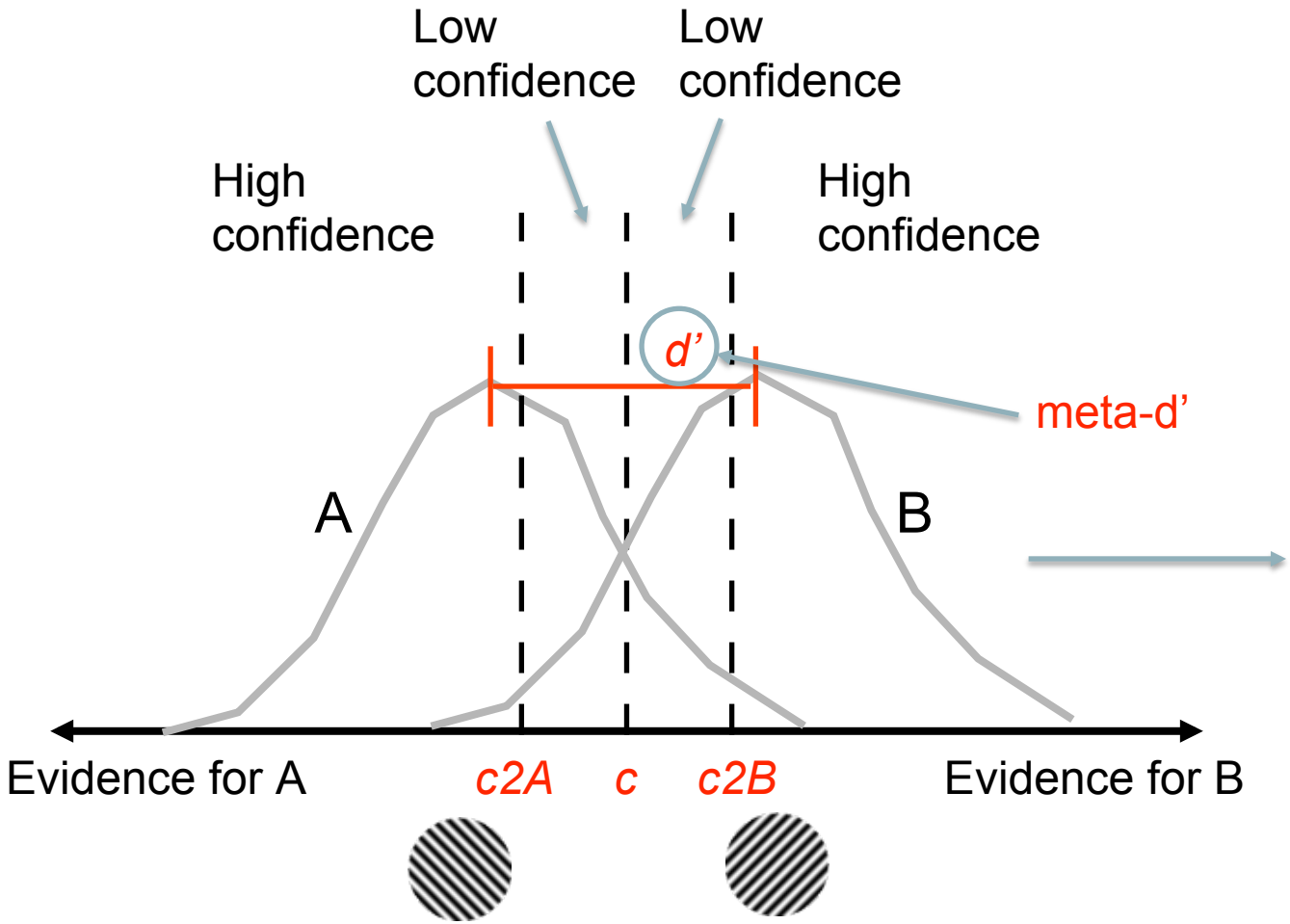
First-order sensitivity



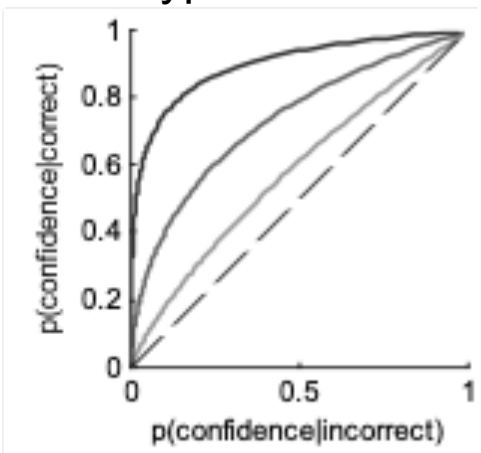


Maniscalco & Lau (2012) *Consciousness & Cognition*; Fleming (2017) *Neuroscience of Consciousness*

Quantifying metacognition (3) - meta-d’



Type 2 ROC

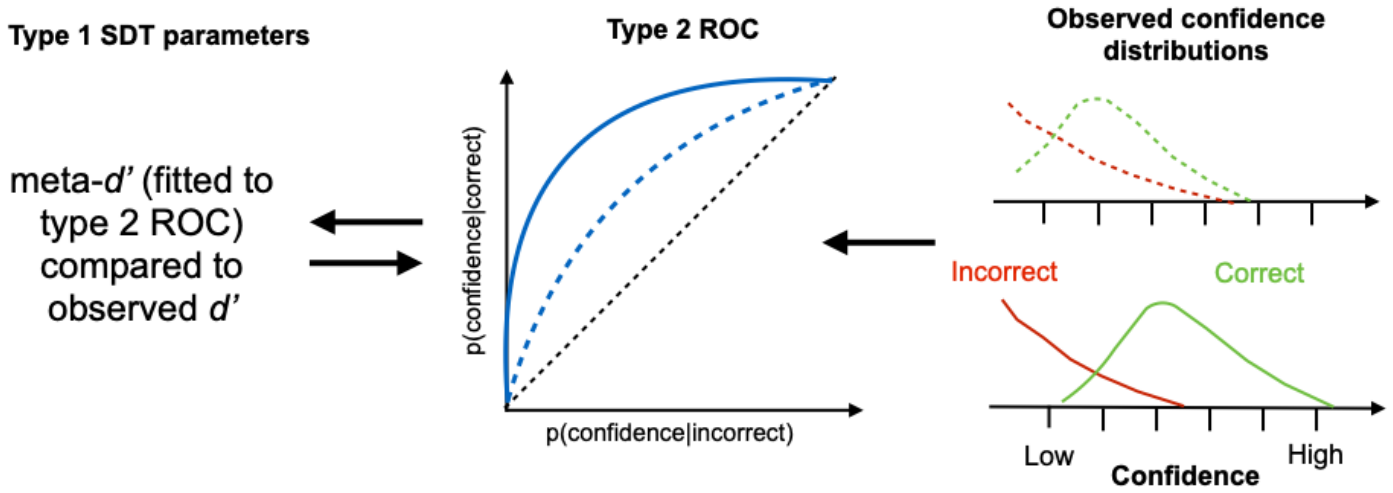


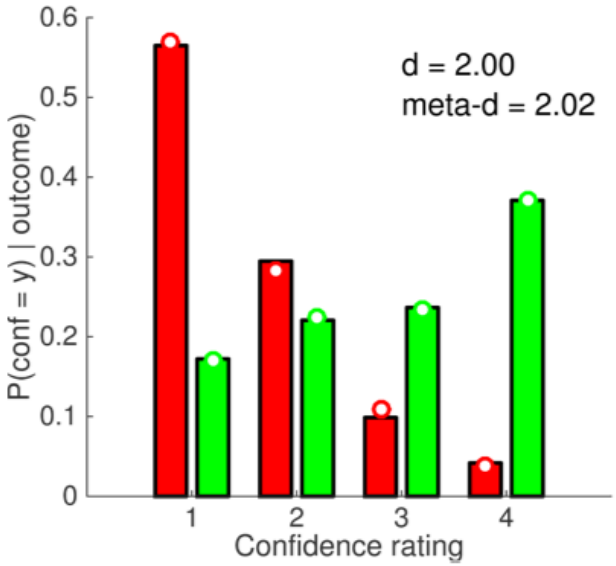
Find parameter set that best fits subjects’ type 2 ROC

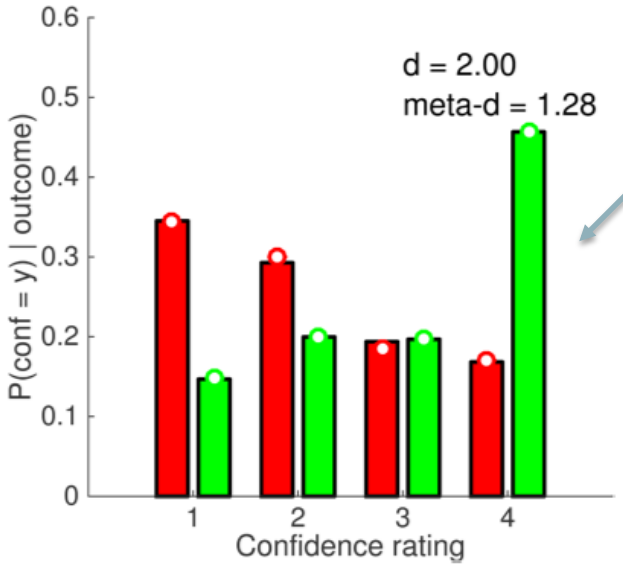
The area under each segment of the curve gives a probability of using a given confidence level

Green & Swets (1966); Maniscalco & Lau (2012) *Consciousness and Cognition*

Quantifying metacognition (3) - meta-d’







Gaussian noise added to confidence ratings

meta-d’/d’ = metacognitive efficiency

Maniscalco & Lau (2012) *Consciousness & Cognition*; Fleming (2017) *Neuroscience of Consciousness*

Taxonomy of metacognitive measures

- **Metacognitive bias** - changes in confidence level despite matched performance (e.g. mean confidence)

- **Metacognitive sensitivity** - how closely one's confidence ratings discriminate between correct and incorrect judgments (e.g. confidence-accuracy correlation; type 2 ROC)

- **Metacognitive efficiency** - subjects' metacognitive capacity given a particular level of task performance (e.g. meta *d' / d'* = « *Mratio* »)

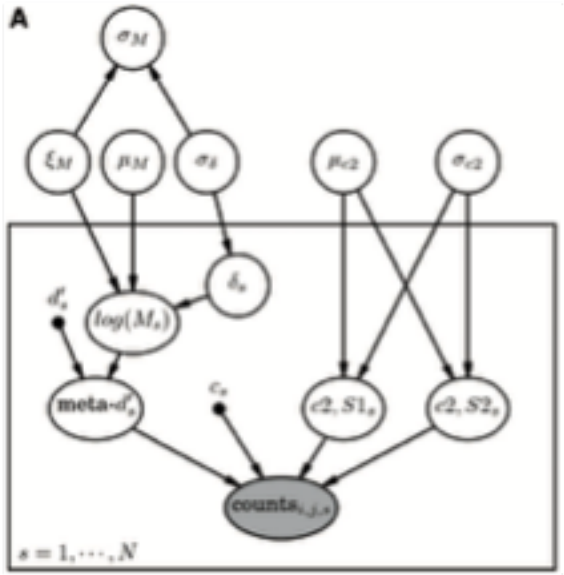
Fleming & Lau (2014) *Frontiers in Neuroscience*

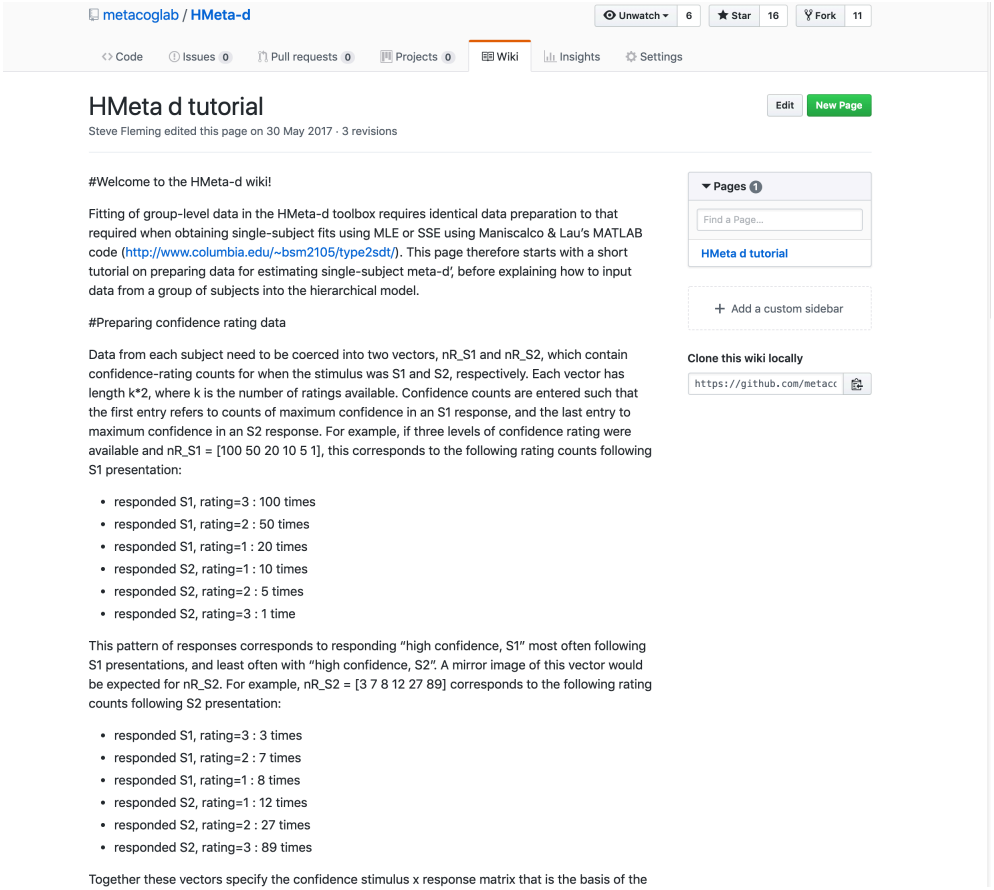
Should I use meta-*d'*?

- **Pros:**
  - Provides principled metric for metacognitive sensitivity in generative model
  - Takes into account both type 1 and type 2 biases
  - Metric is in units of type 1 d', easy to control for performance (e.g. using meta-d'/d')

- **Cons:**
  - Currently only developed for 2-choice discrimination tasks (need to specify 2 x 2 stimulus/response table)
  - Equal-variance Gaussian assumptions may not hold for some tasks
  - Biased estimates with low trial numbers; use HMeta-d!

HMeta-d toolbox





metacoglab / HMeta-d

Unwatch 6 | Star 16 | Fork 11

Code | Issues 0 | Pull requests 0 | Projects 0 | Wiki | Insights | Settings

HMeta d tutorial

Edit | New Page

Steve Fleming edited this page on 30 May 2017 · 3 revisions

#Welcome to the HMeta-d wiki!

Fitting of group-level data in the HMeta-d toolbox requires identical data preparation to that required when obtaining single-subject fits using MLE or SSE using Maniscalco & Lau's MATLAB code (http://www.columbia.edu/~bsm2105/type2sdt/). This page therefore starts with a short tutorial on preparing data for estimating single-subject meta-d', before explaining how to input data from a group of subjects into the hierarchical model.

#Preparing confidence rating data

Data from each subject need to be coerced into two vectors, nR_S1 and nR_S2, which contain confidence-rating counts for when the stimulus was S1 and S2, respectively. Each vector has length k*2, where k is the number of ratings available. Confidence counts are entered such that the first entry refers to counts of maximum confidence in an S1 response, and the last entry to maximum confidence in an S2 response. For example, if three levels of confidence rating were available and nR_S1 = [100 50 20 10 5 1], this corresponds to the following rating counts following S1 presentation:

- responded S1, rating=3 : 100 times
- responded S1, rating=2 : 50 times
- responded S1, rating=1 : 20 times
- responded S2, rating=1 : 10 times
- responded S2, rating=2 : 5 times
- responded S2, rating=3 : 1 time

This pattern of responses corresponds to responding "high confidence, S1" most often following S1 presentations, and least often with "high confidence, S2". A mirror image of this vector would be expected for nR_S2. For example, nR_S2 = [3 7 8 12 27 89] corresponds to the following rating counts following S2 presentation:

- responded S1, rating=3 : 3 times
- responded S1, rating=2 : 7 times
- responded S1, rating=1 : 8 times
- responded S2, rating=1 : 12 times
- responded S2, rating=2 : 27 times
- responded S2, rating=3 : 89 times

Together these vectors specify the confidence stimulus x response matrix that is the basis of the

Pages 1

Find a Page...

HMeta d tutorial

+ Add a custom sidebar

Clone this wiki locally

https://github.com/metacc

https://github.com/metacoglab/HMeta-d

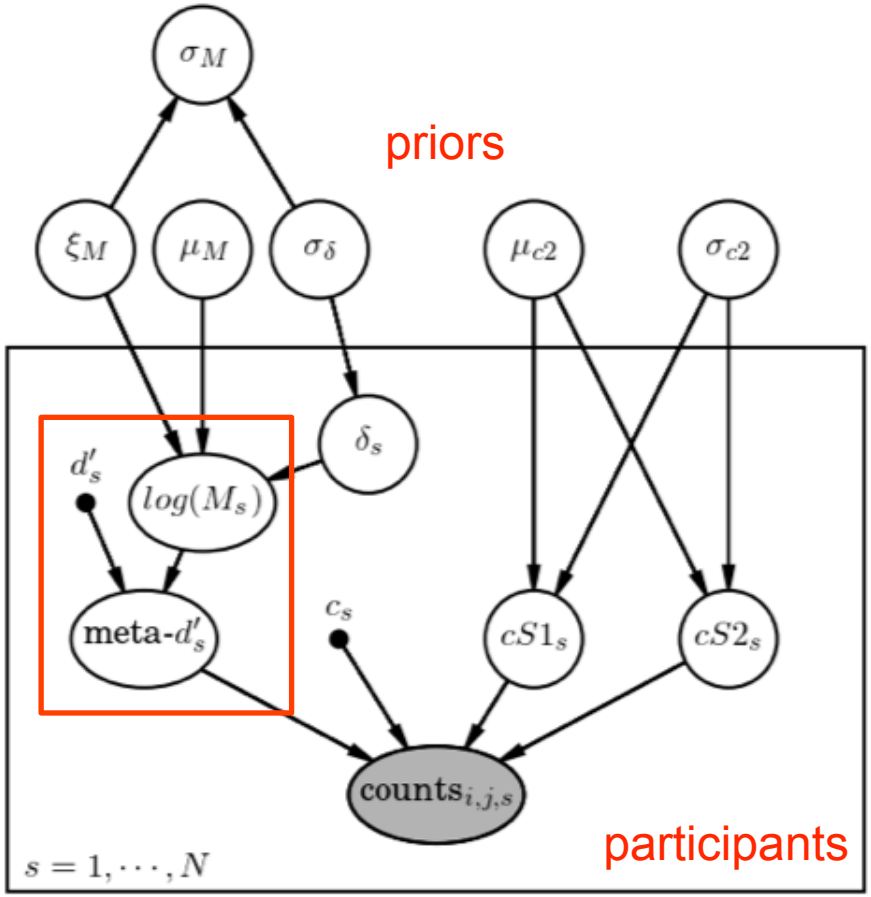
Fleming (2017) *Neuroscience of Consciousness*

Advantages of hierarchical approach

1. Point estimates of meta-d' are noisy, particularly with small numbers of trials; frequentist estimates of hit/false alarm rates fail to account for uncertainty in these rates
2. A hierarchical Bayesian approach is the correct way to combine information about within- and between-subject uncertainty, each subject mutually constrains the group fit
3. When fitting SDT models to data, padding (edge correction) is often applied to avoid zero cell counts when no all types of responses are present; generative multinomial model avoids this
4. Testing group-level hypotheses is straightforward. E.g. can directly compare posterior distribution over metacognitive sensitivity for patients and controls

Fleming (2017) *Neuroscience of Consciousness*

Hierarchical model for meta-*d'* (HMeta-d)



$$\mu_{c2} \sim \mathcal{N}(0, 10)$$

$$\sigma_{c2} \sim \mathcal{HN}(10)$$

$$\mu_M \sim N(0, 1)$$

$$\sigma_M = |\xi_M| \times \delta_s$$

$$\xi_M \sim Beta(1,1)$$

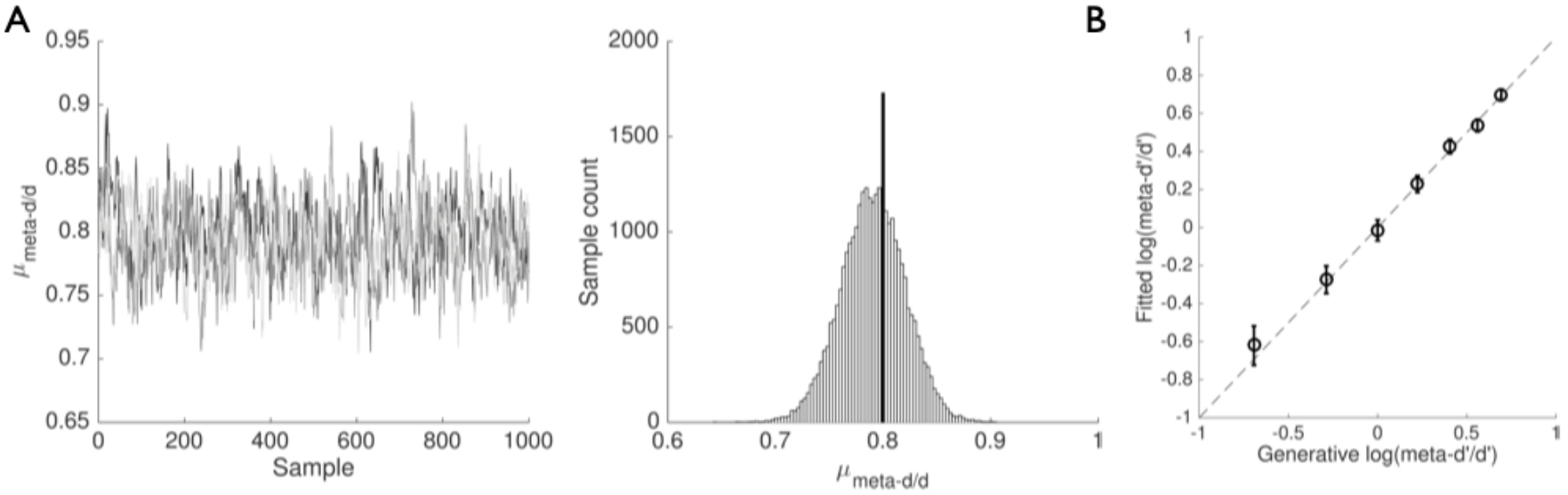
$$\sigma_\delta \sim \mathcal{HN}(1)$$

Fleming (2017) *Neuroscience of Consciousness*

Hierarchical model for meta-*d'* (HMeta-d)

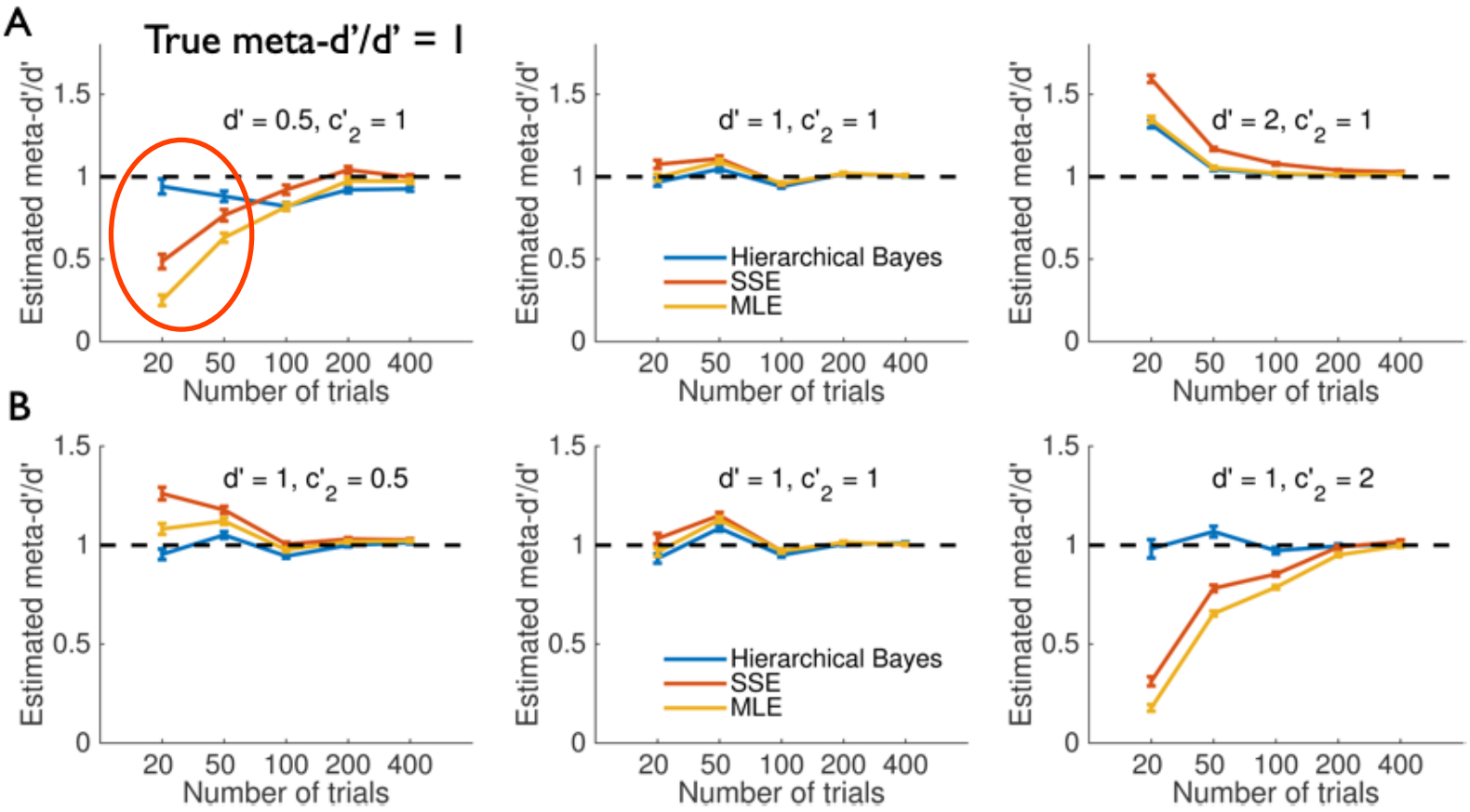
https://github.com/smfleming/HMeta-d

MCMC samples of group-level metacognitive efficiency:



Fleming (2017) *Neuroscience of Consciousness*

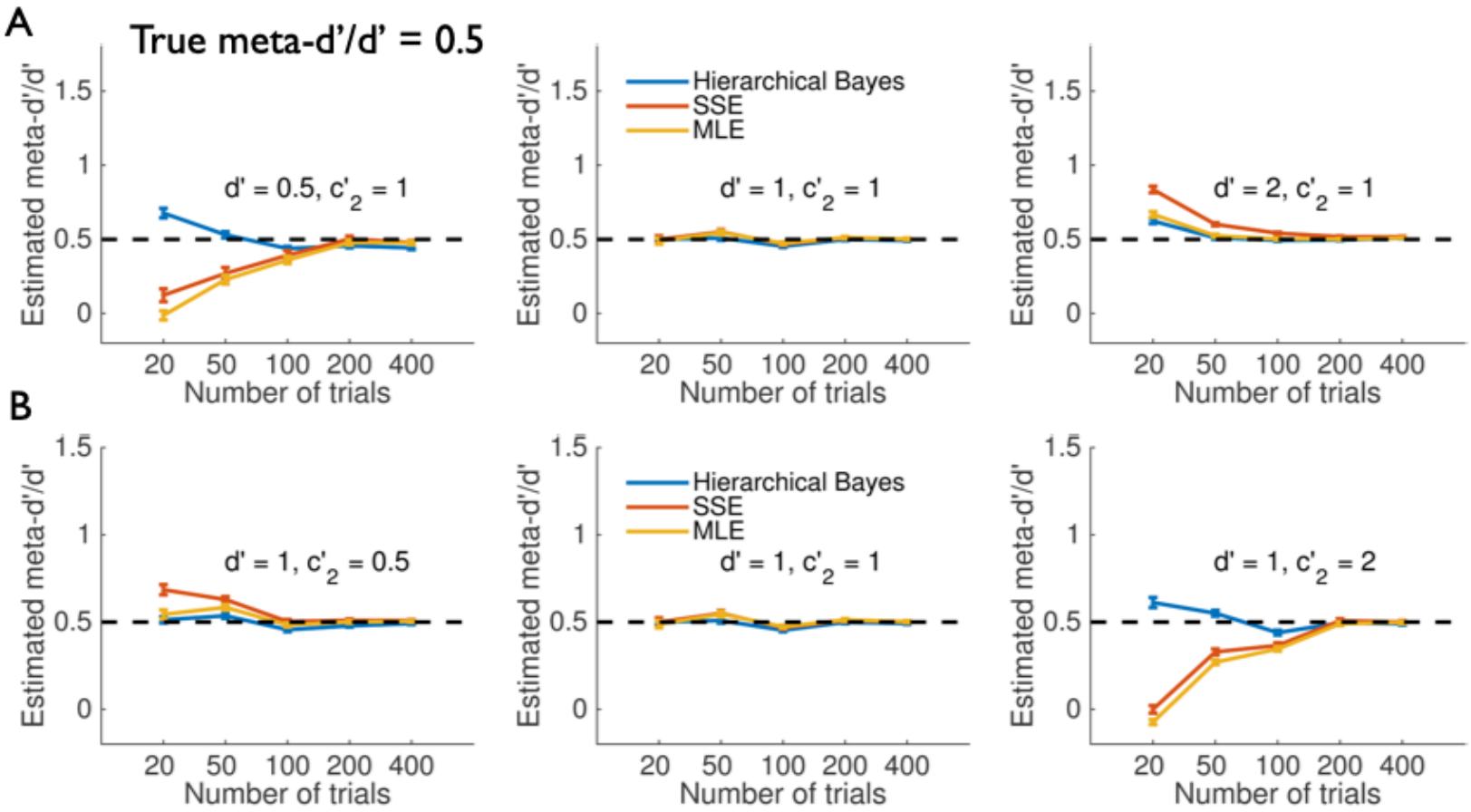
Hierarchical model for meta-*d'* (HMeta-d)



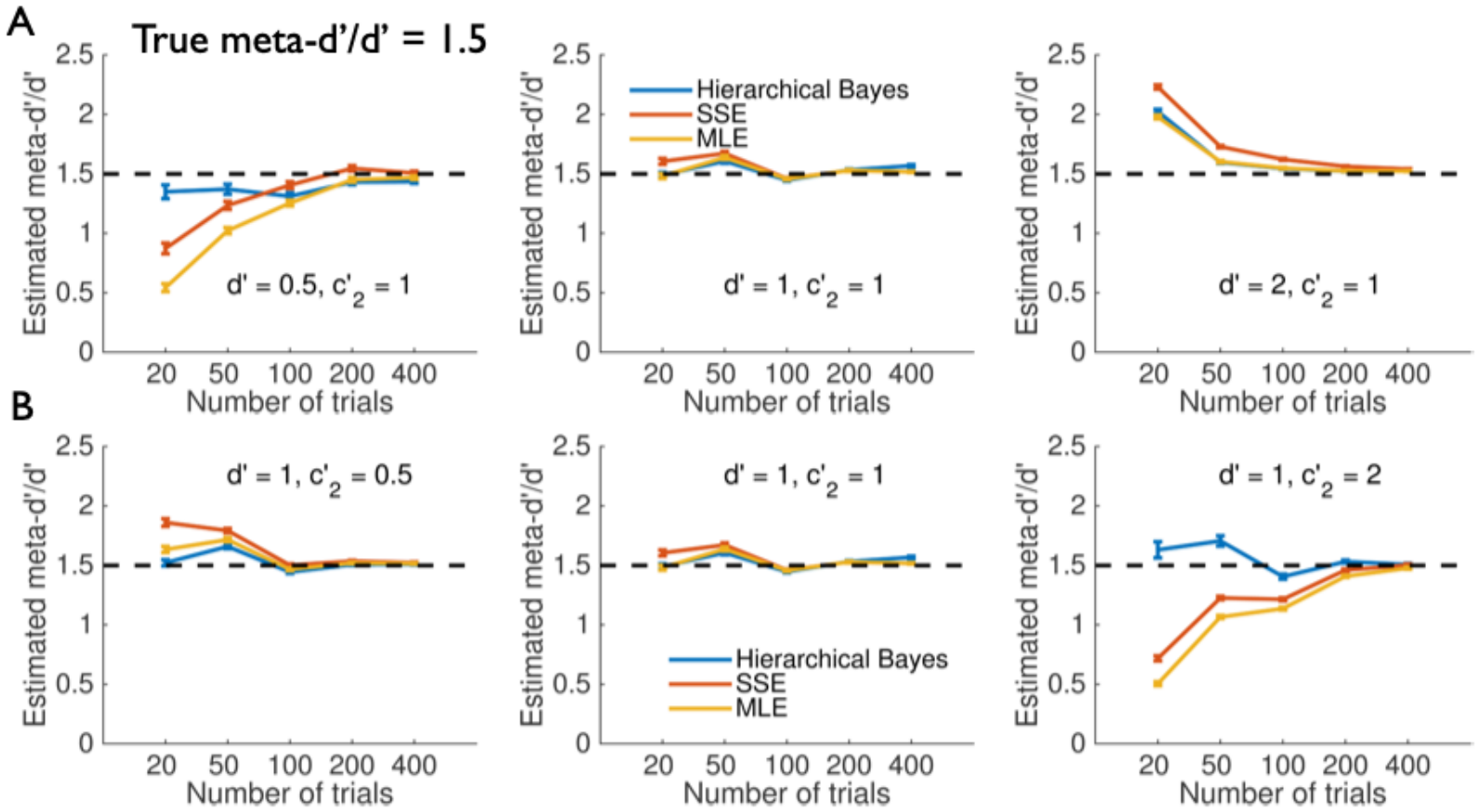
Point-estimate approaches underestimate metacognitive efficiency when (type 1) d' is low

Why is HMeta-d better? Shrinkage to prior OR capitalises on hierarchy across participants…

Fleming (2017) *Neuroscience of Consciousness*

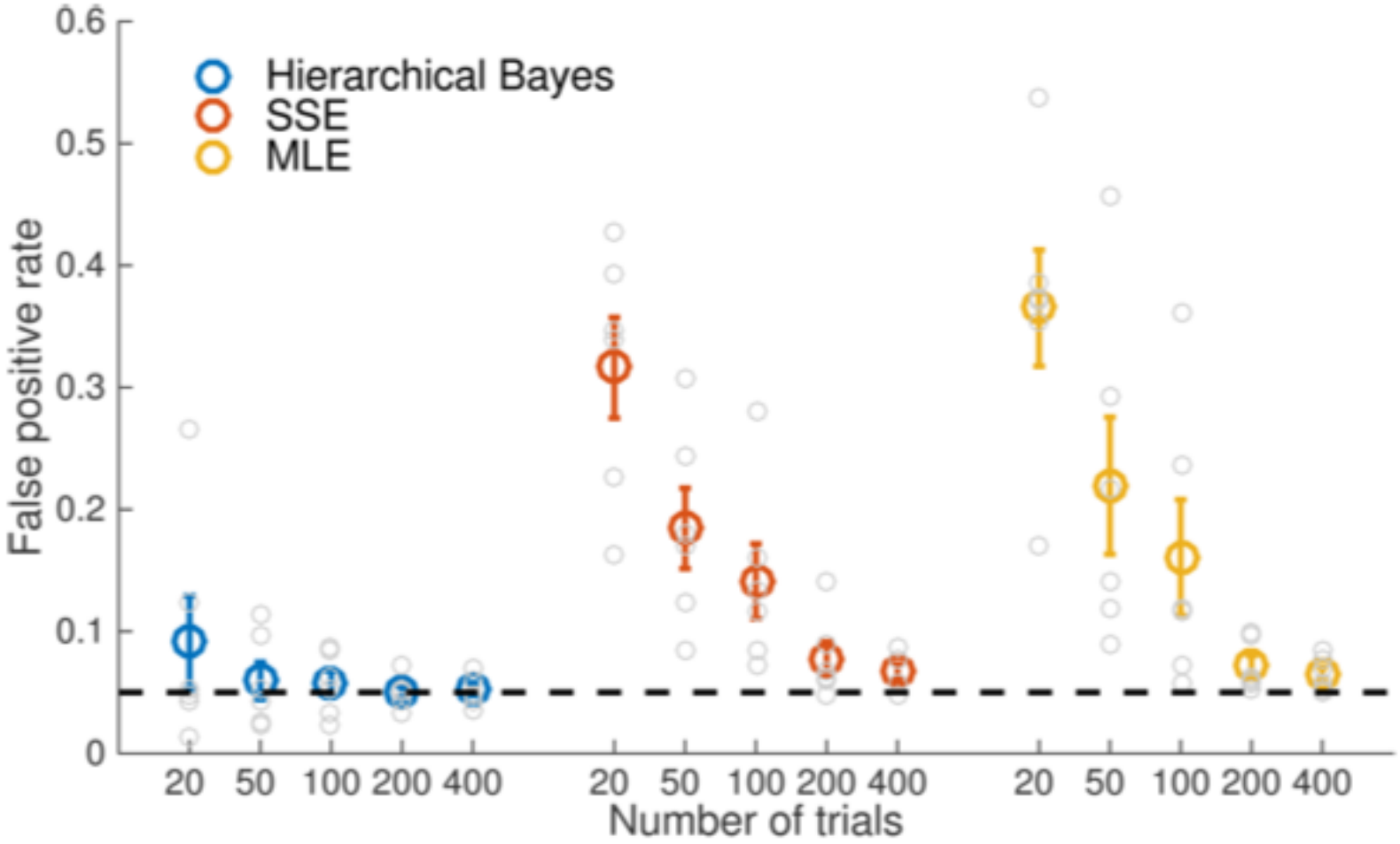


Fleming (2017) *Neuroscience of Consciousness*



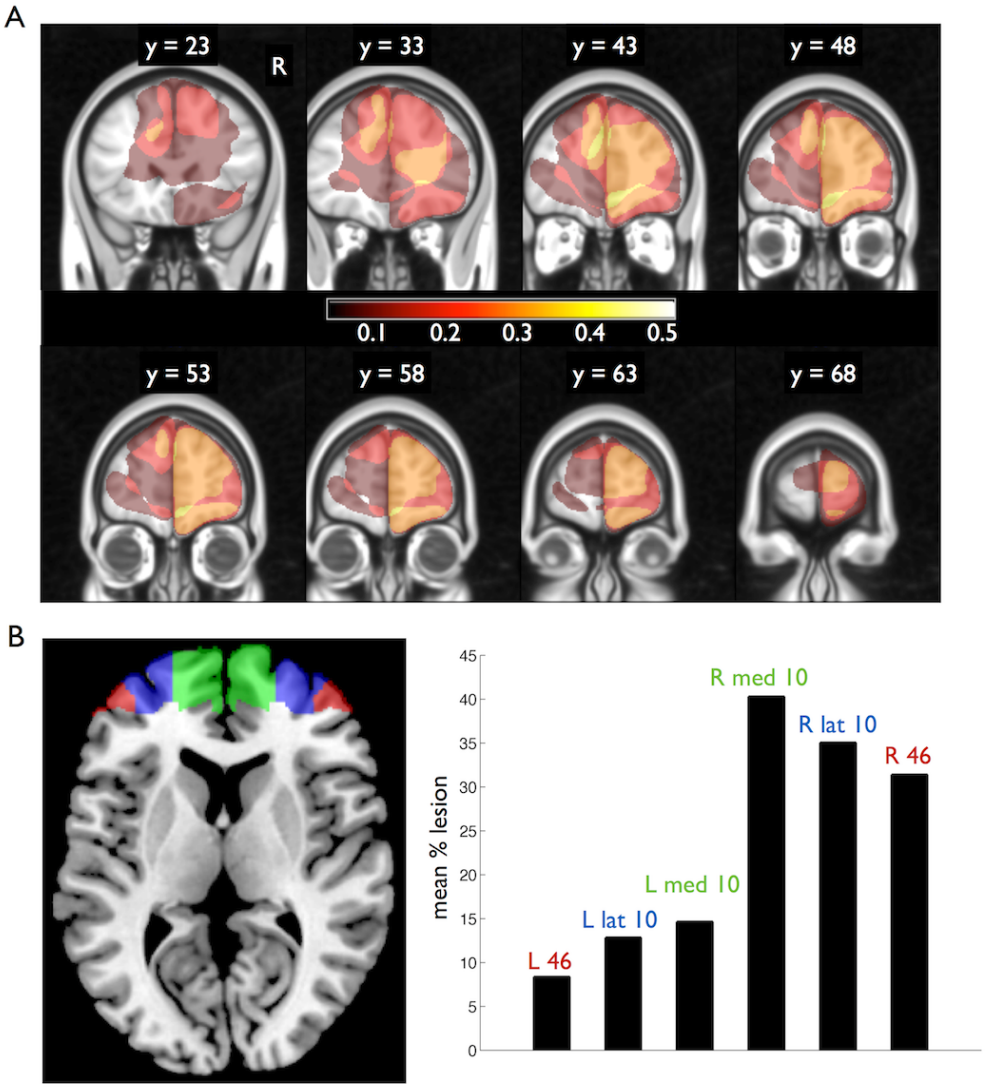
Fleming (2017) *Neuroscience of Consciousness*

Hierarchical model for meta-*d'* (HMeta-d)

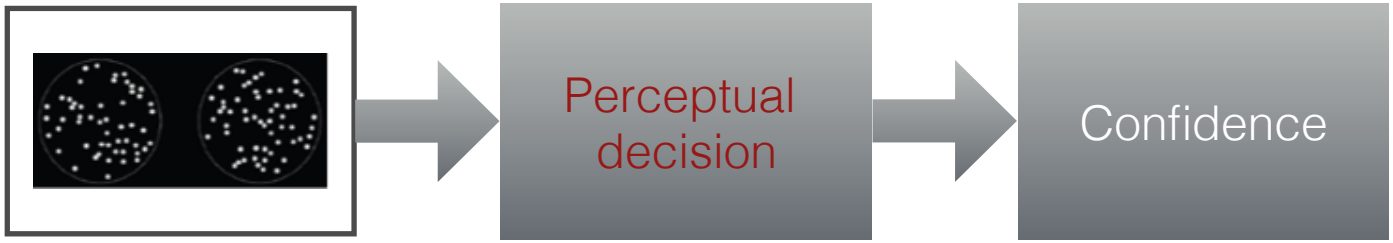


Fleming (2017) *Neuroscience of Consciousness*

Empirical example



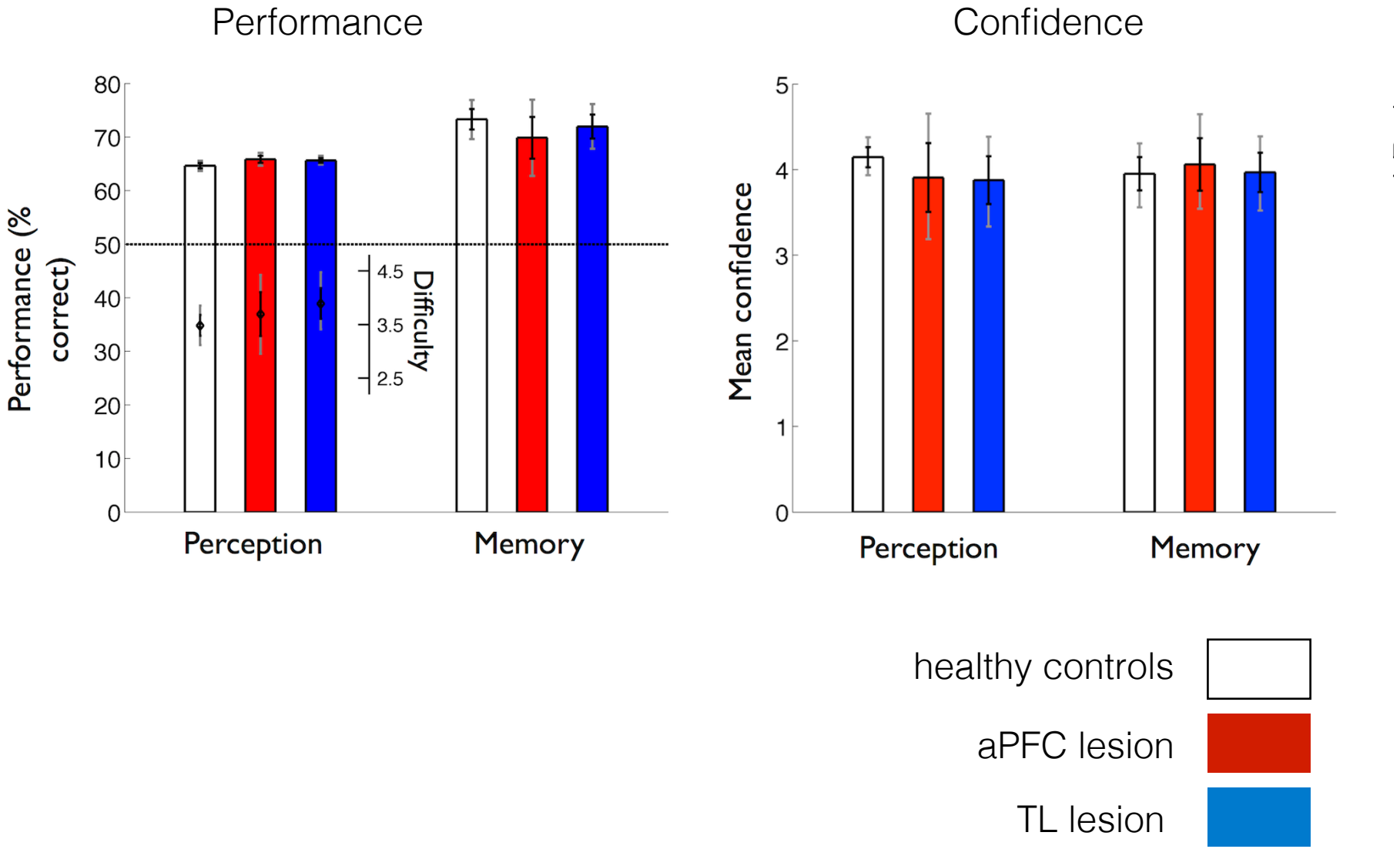
Fleming, Ryu, Golfinos & Blackmon *Brain* (2014)



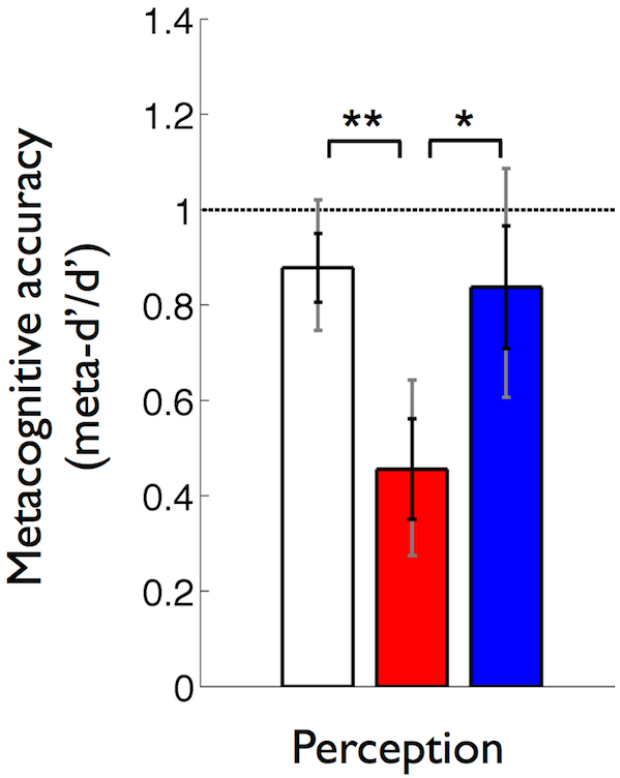
More dots, L or R?

Old, L or R?

Fleming, Ryu, Golfinos & Blackmon ***Brain*** (2014)



Fleming, Ryu, Golfinos & Blackmon ***Brain*** (2014)



healthy controls

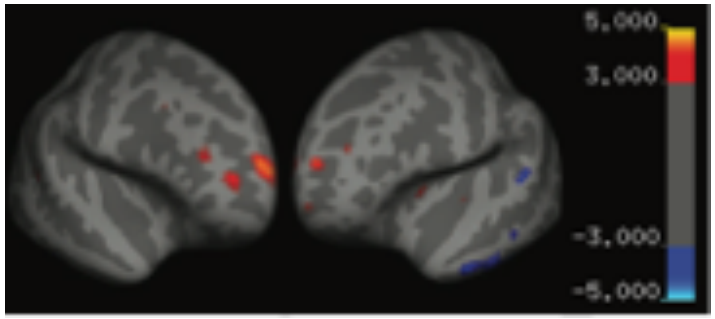
aPFC lesion

TL lesion

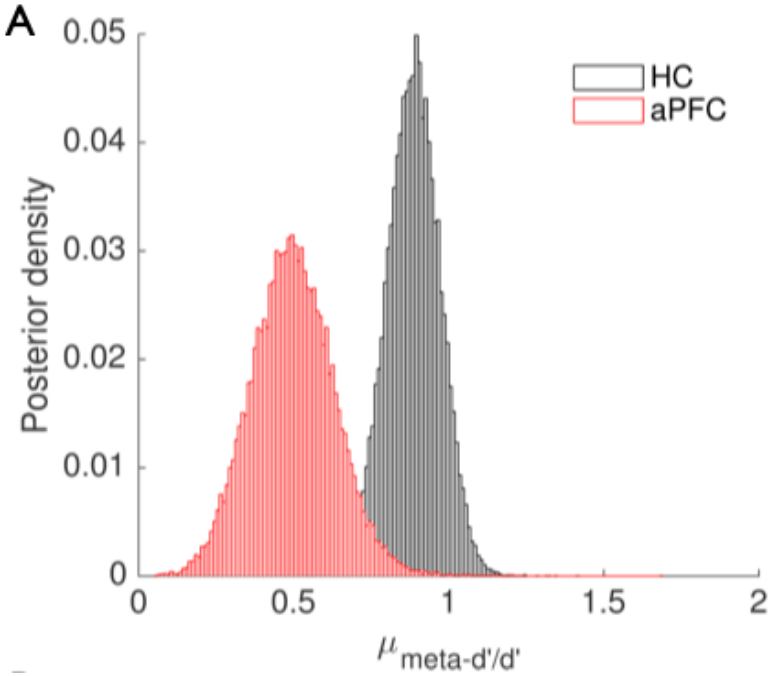
Why no metacognitive deficit for memory following aPFC lesion?
Redundancy? Reorganisation? Intact parietal cortex may compensate?

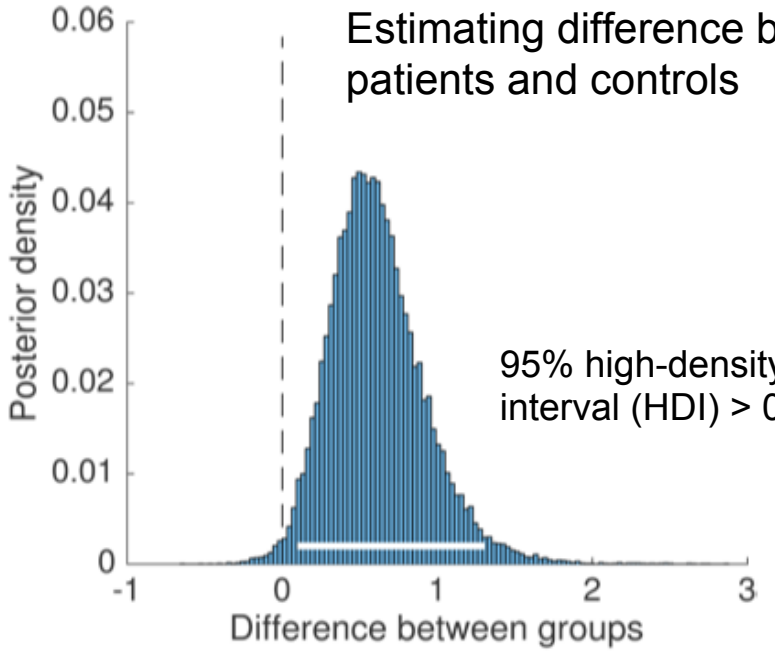
Fleming, Ryu, Golfinos & Blackmon ***Brain*** (2014)

Re-analysis of perceptual confidence using HMeta-d



Fleming et al. (2010) *Science*





Estimating difference between patients and controls

95% high-density interval (HDI) > 0

Fleming (2017) *Neuroscience of Consciousness*

Metacognition and psychopathology

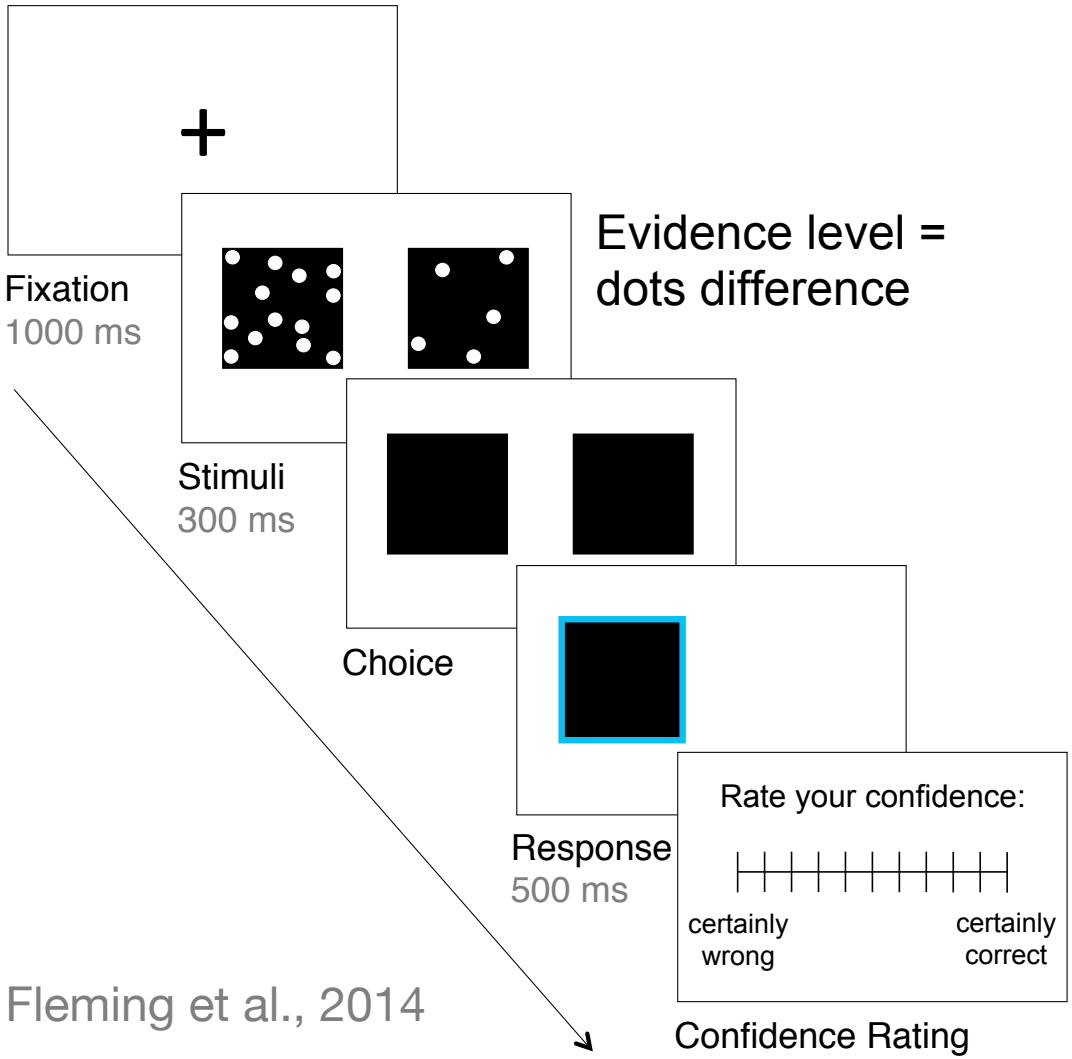
A ‘general population’ sample
Total N=995 participants

Perceptual decision-making task

Self-reported symptom questionnaires

Rouault*, Seow*, Gillan and Fleming (2017) *Biological Psychiatry*

Perceptual decision-making task



Fleming et al., 2014

QUANTIFY

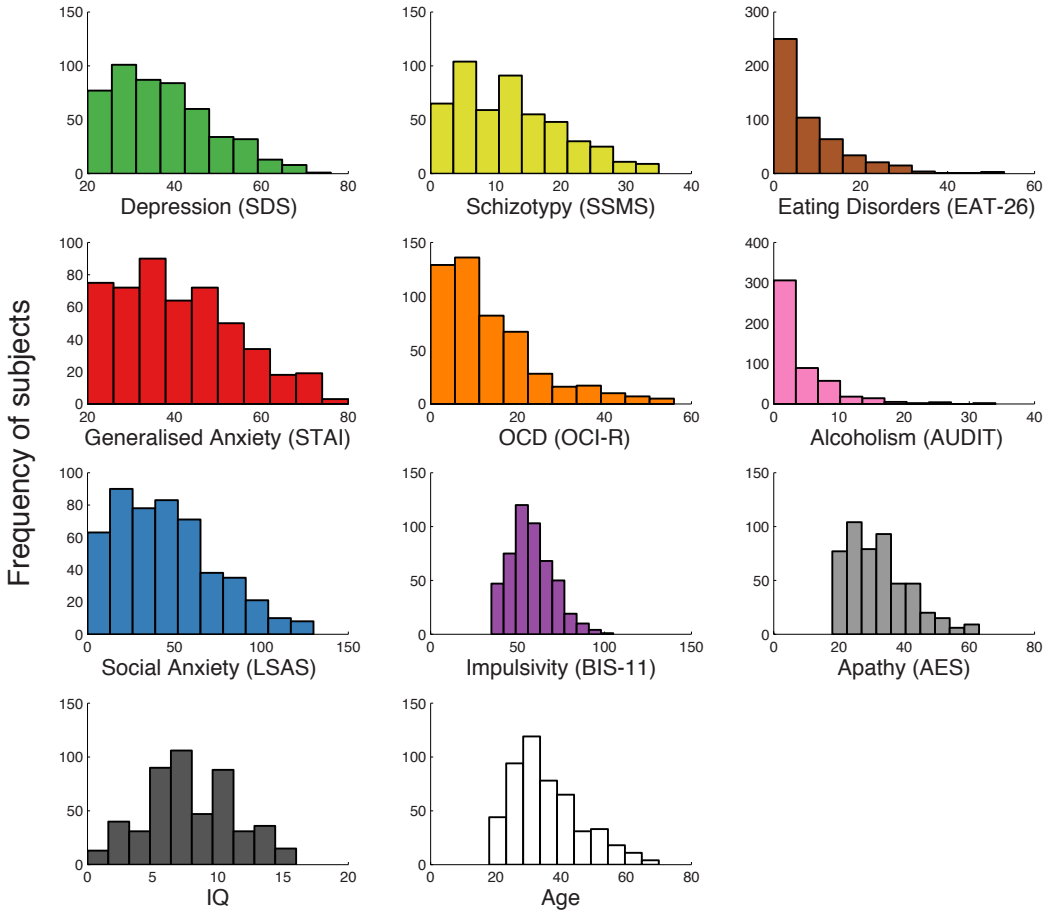
- **Decision process**

Accuracy
(Drift-diffusion model)

- **Metacognition**

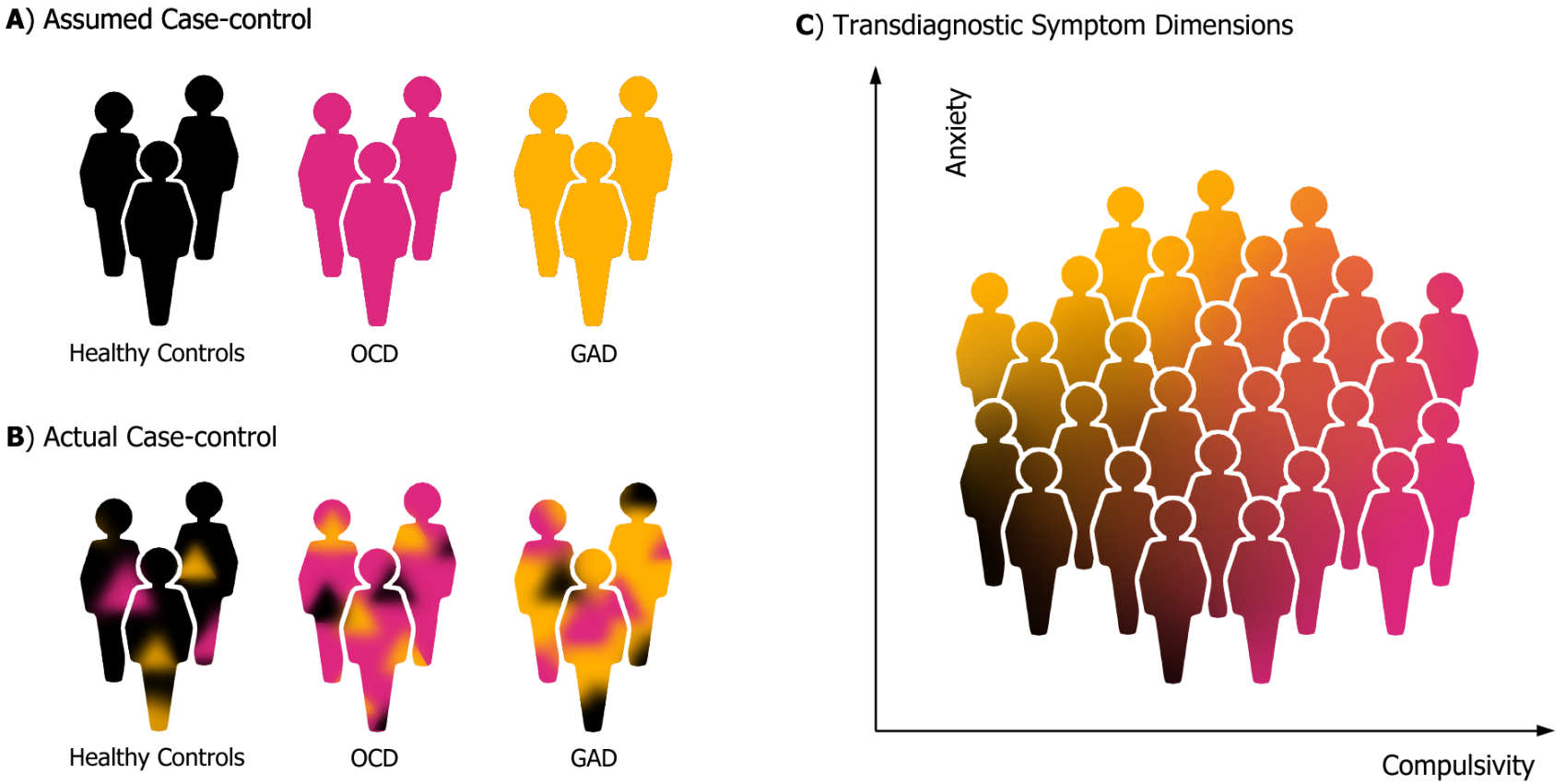
⇒ **Confidence level**
⇒ **Metacognitive efficiency**

Self-reported psychiatric symptoms



Strong **correlations** between individual questionnaire scores, consistent with **comorbidities** between diagnostic categories

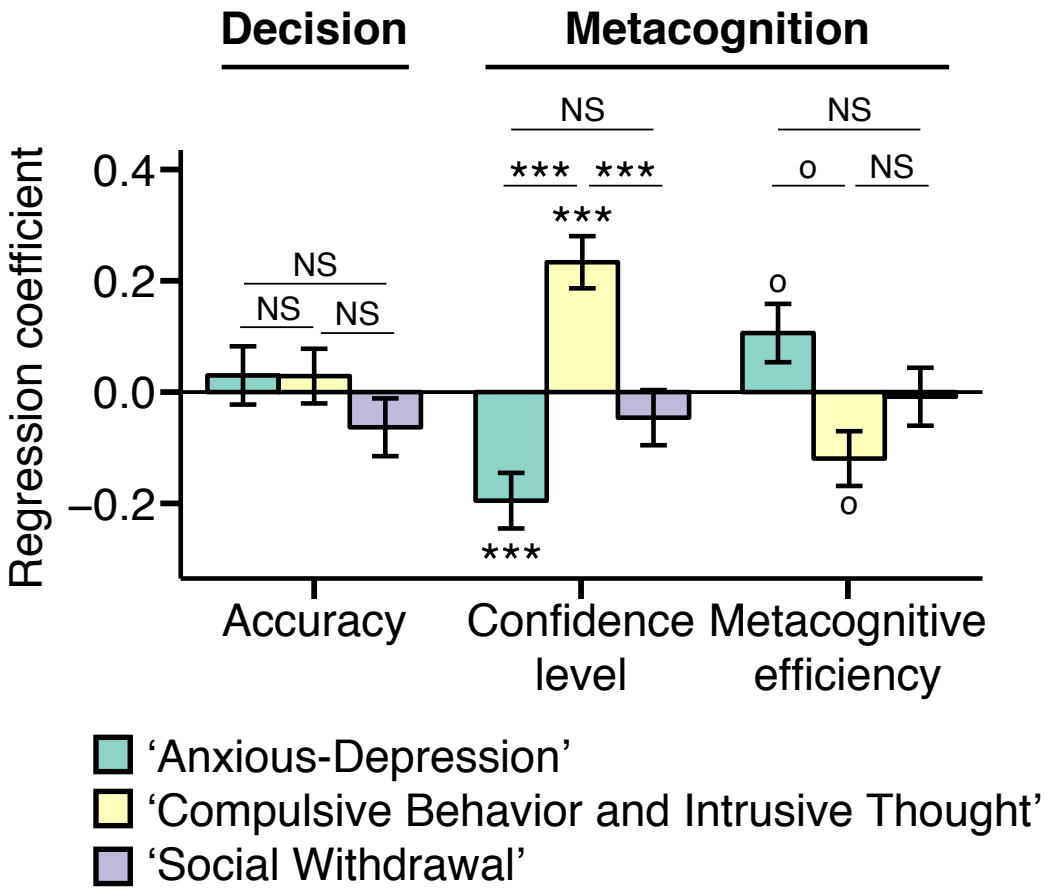
A transdiagnostic approach



Strong **correlations** between individual questionnaire scores, consistent with **comorbidities** between diagnostic categories

Seow*, Rouault*, Gillan and Fleming (2021) *Biological Psychiatry*

Inter-individual variability in metacognition



Anxious/Depression

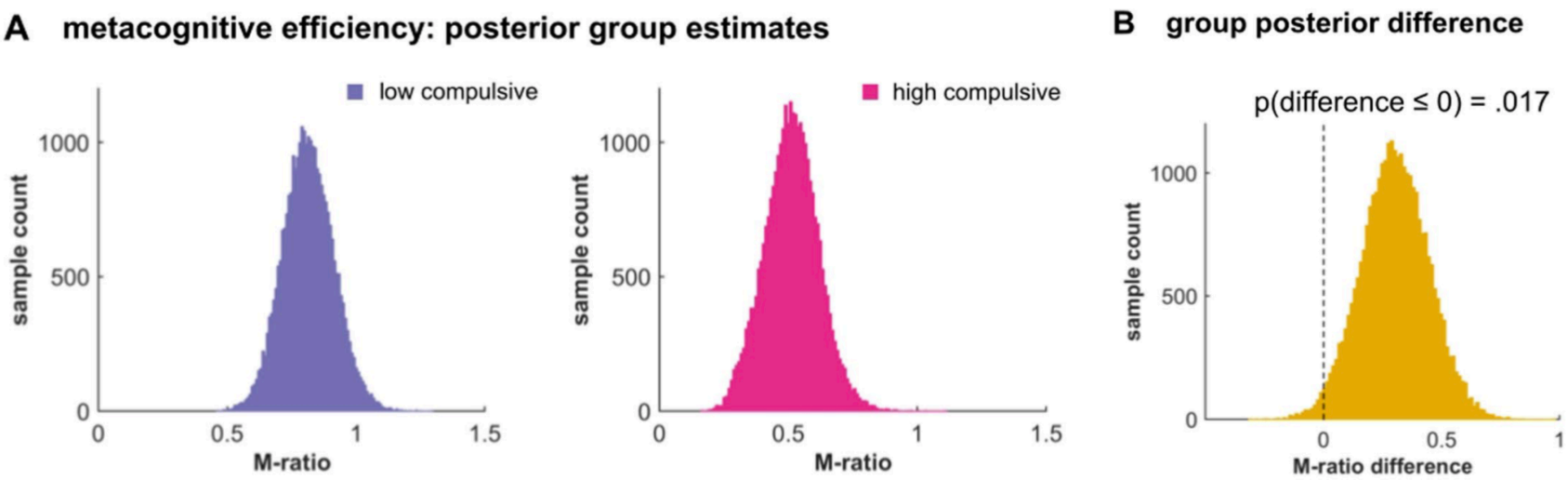
Confidence ↓
Metacognitive efficiency ↑

Compulsive/Intrusive

Confidence ↑
Metacognitive efficiency ↓

Rouault*, Seow*, Gillan and Fleming (2017) *Biological Psychiatry*

Metacognition and computational psychiatry



Reduced perceptual metacognitive efficiency in individuals with high compulsivity (N=20 per group)

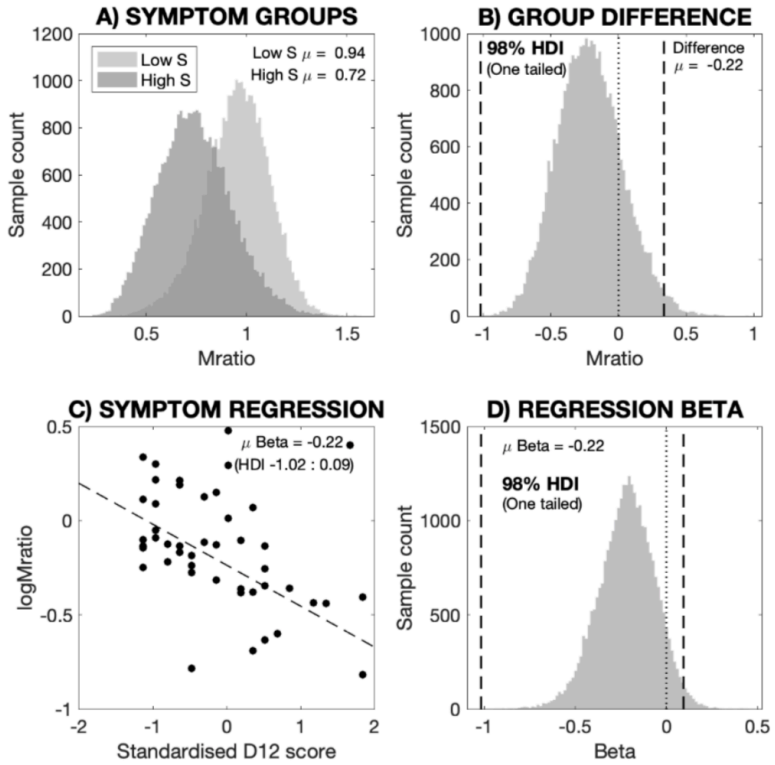
Hauser et al., (2017) *Scientific Reports*

« Psychiatry-informed » prior

Can we embed trait psychopathology directly into metacognitive efficiency estimation?

Hypothesis: trait characteristics will impact metacognitive behaviour: include as a co-regressor

- Comparison of high and low symptom groups (asthma) in an interoception breathing task:



Moutoussis et al., (2019) *Frontiers*; Harrison et al., (2020) ***Biological Psychology***

Beyond the meta-*d'* framework

Ongoing efforts to refine estimates of metacognitive ability

- Addition of metacognitive noise over sensory noise: "ReMeta toolbox":

✓ The model assumes that confidence results from a continuous but noisy and potentially biased transformation of decision values, described by a confidence link function.

✓ A canonical set of metacognitive noise distributions is introduced which differ, amongst others, in their predictions about metacognitive sign flips of type 1 decision values.

✓ Metacognitive noise and bias parameters correlate with conventional behavioral measures.
✓ But in contrast to conventional measures, metacognitive noise parameters inferred from the model are shown to be independent of type 1 performance.

- Dependence on response times (Desender et al., 2022 Nat Commun)

Guggenmos (2021) *BioRxiv. Reverse engineering of metacognition. ReMeta toolbox*

Take home messages: models of metacognition

- There is not one method *systematically* better than the other for measuring metacognition
- Each method has pros and cons depending on the nature of your empirical data
- This is currently an active area of research: e.g. continuous refinement of fitting techniques

Thank you for your attention