Models of metacognition

Computational Psychiatry Course – Zurich 2024

Dr. Marion ROUAULT

(marion.rouault@gmail.com)

Centre National de la Recherche Scientifique & Paris Brain Institute

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

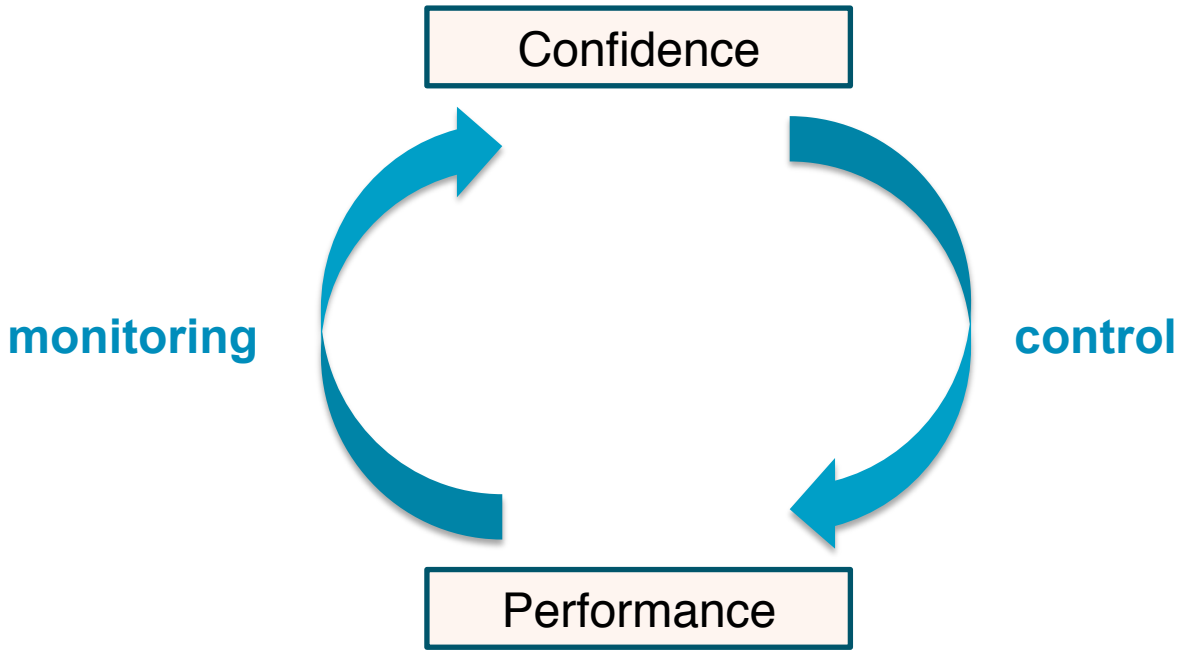
- Most perceptions, memories and choices are accompanied by subjective estimates of their reliability i.e. **confidence** estimates

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

- **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?

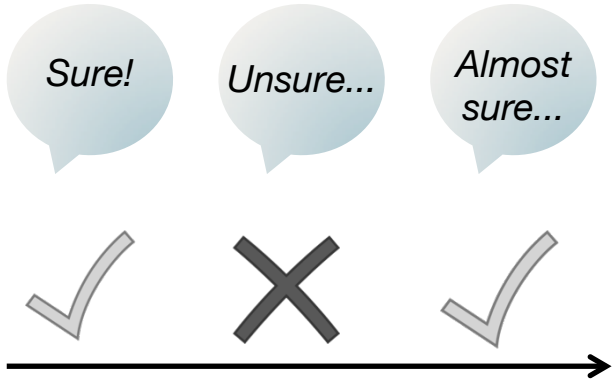
Inaccurate metacognitive knowledge of cognitive and physical impairments is common in **psychiatric** and **neurological** disorders and in **healthy aging**

Insight = the capacity for accurate metacognition

Metacognition: a central aspect of neurological and psychiatric pathologies

Metacognition in the lab

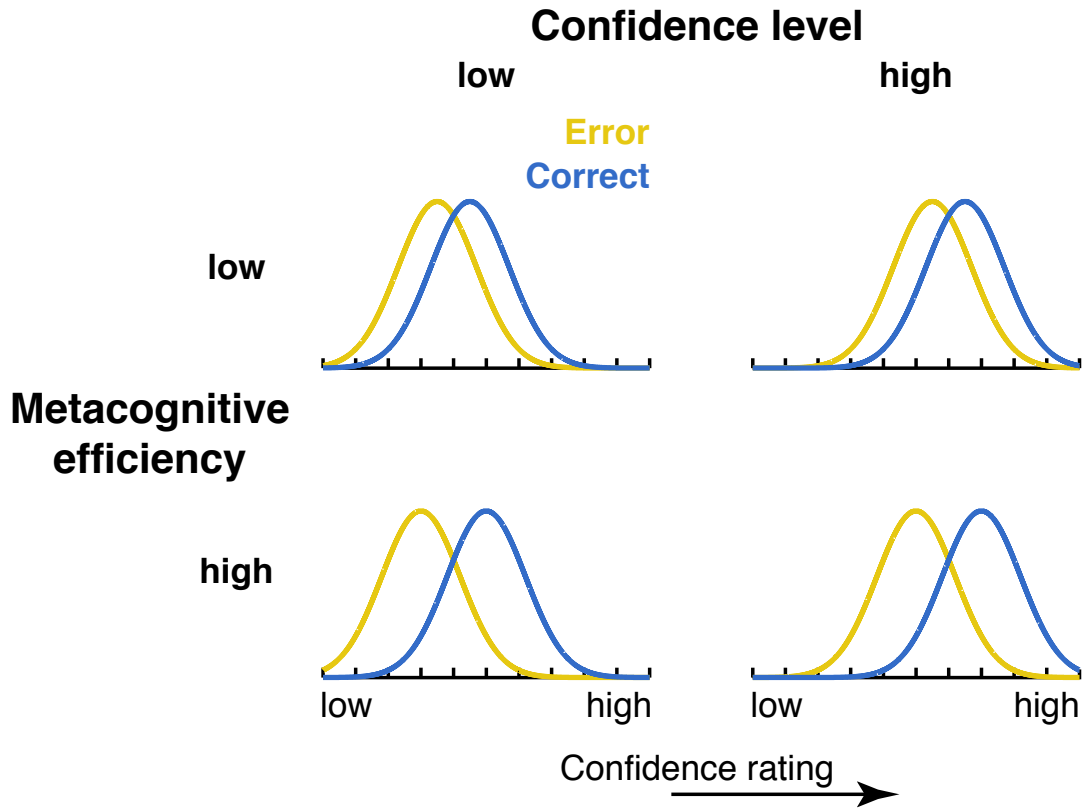
- Not possible to assess metacognition from a single judgment
- Need multiple judgments over time, examine **statistical association** between behavioural responses and metacognitive judgments



	High confidence	Low confidence
✓ Correct		
✗ Incorrect		

High metacognitive ability
Low metacognitive ability

Two independent metrics of metacognition

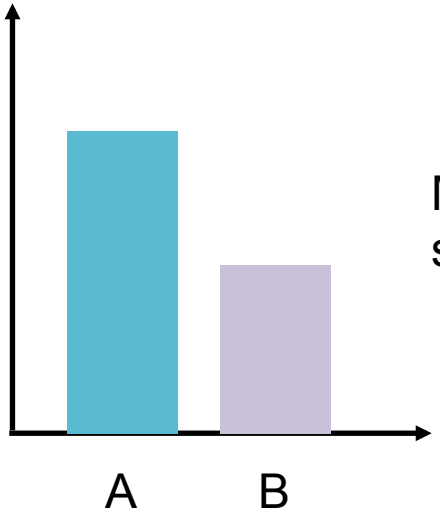


Other terminology in the literature:
-Confidence level: calibration, metacognitive bias, self-perceived ability
-Efficiency: sensitivity, discrimination, resolution, metacognitive awareness, insight

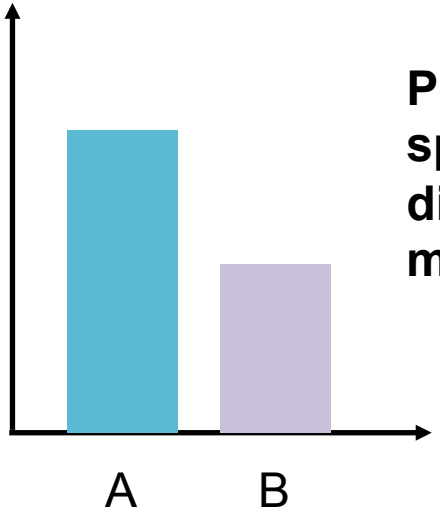
Fleming & Lau (2014) *Frontiers*

Importance of taking into account performance

First-order performance

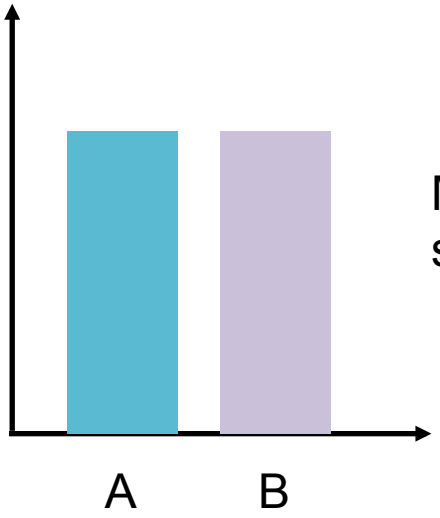


Metacognitive sensitivity

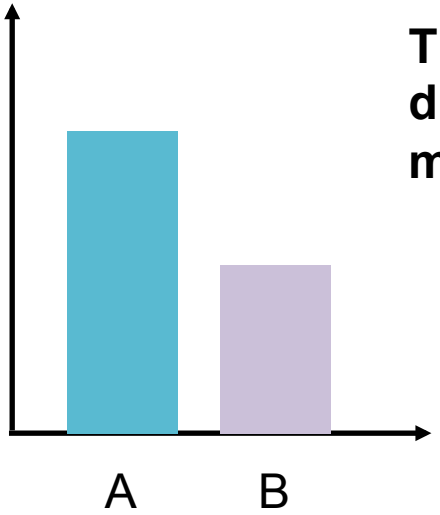


Possibly spurious difference in metacognition

First-order performance



Metacognitive sensitivity



True difference in metacognition

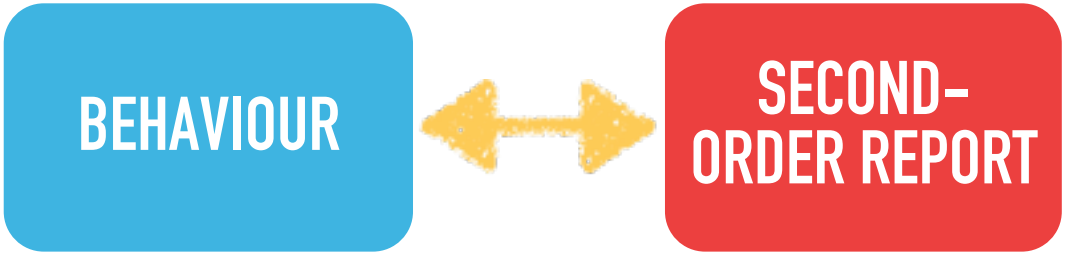
How to measure metacognitive sensitivity?

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…]

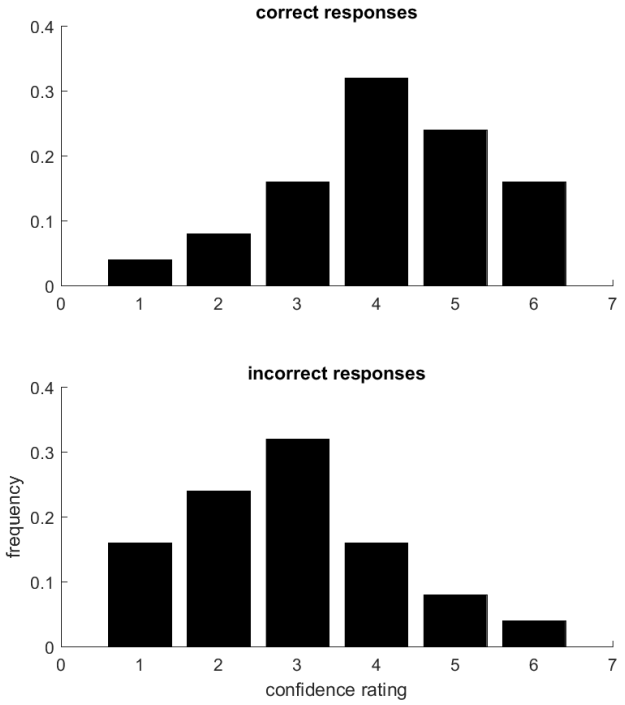
Quantifying metacognition (2) - AUROC2

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

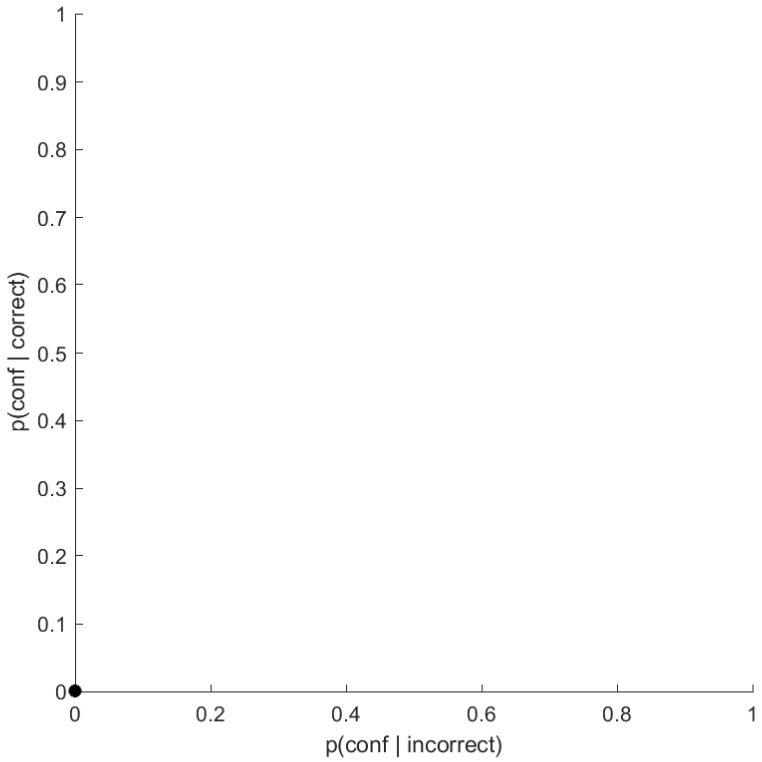
INTUITION: the more distinct the confidence distributions for correct and for incorrect responses are, the more insight one has into the quality of individual decisions

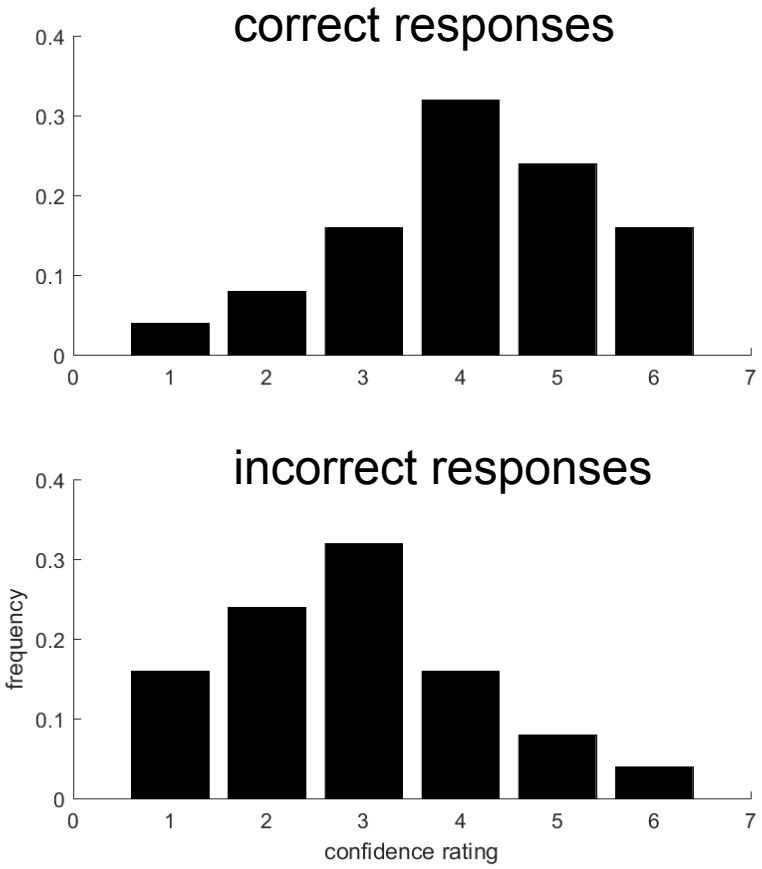


with thanks to Matan Mazor

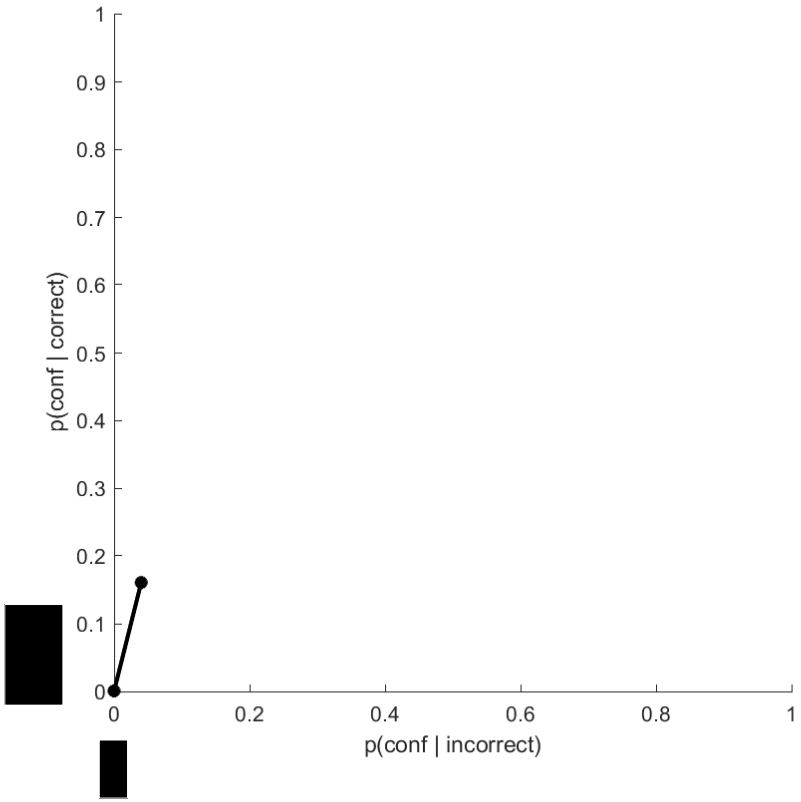
Clarke, Birdsall & Tanner (1959) *J Acoust Soc Am*; Galvin et al. (2003) *Psych Bull Rev*

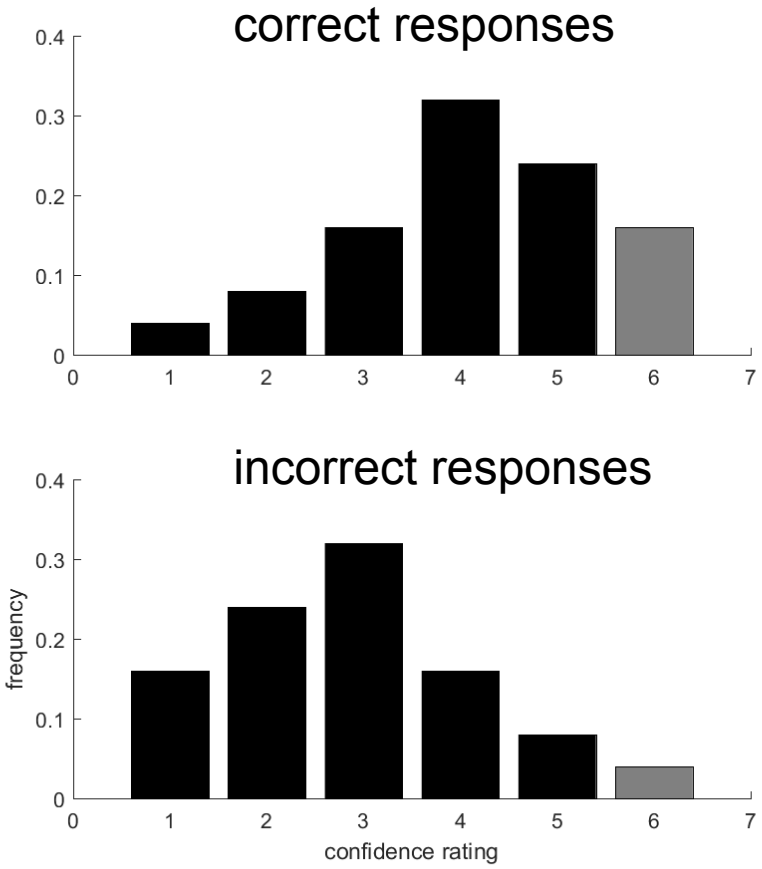
Type 2 ROCs


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

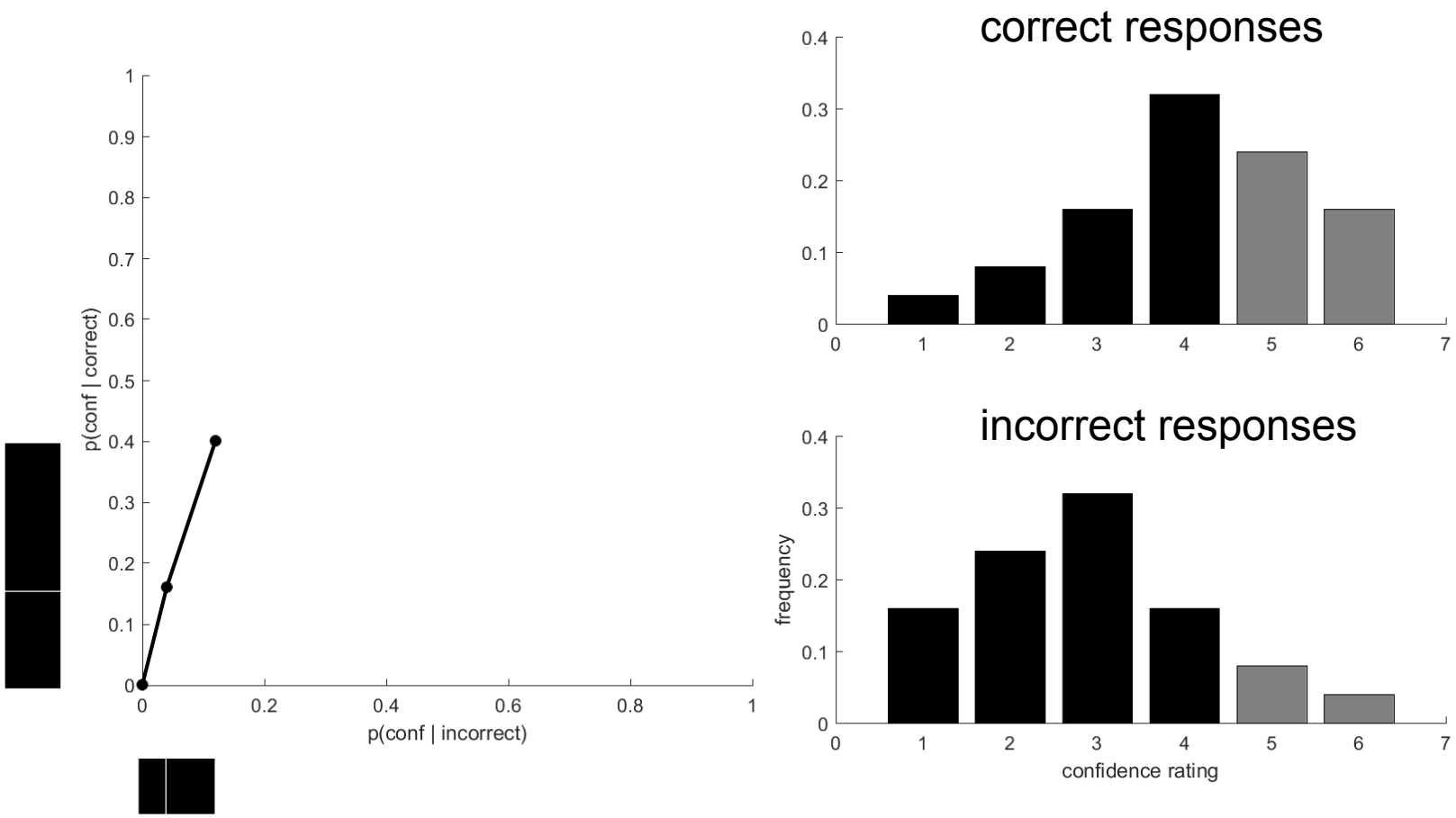

correct responses
incorrect responses
frequency
confidence rating
0
0.1
0.2
0.3
0.4
0
1
2
3
4
5
6
7

Type 2 ROCs

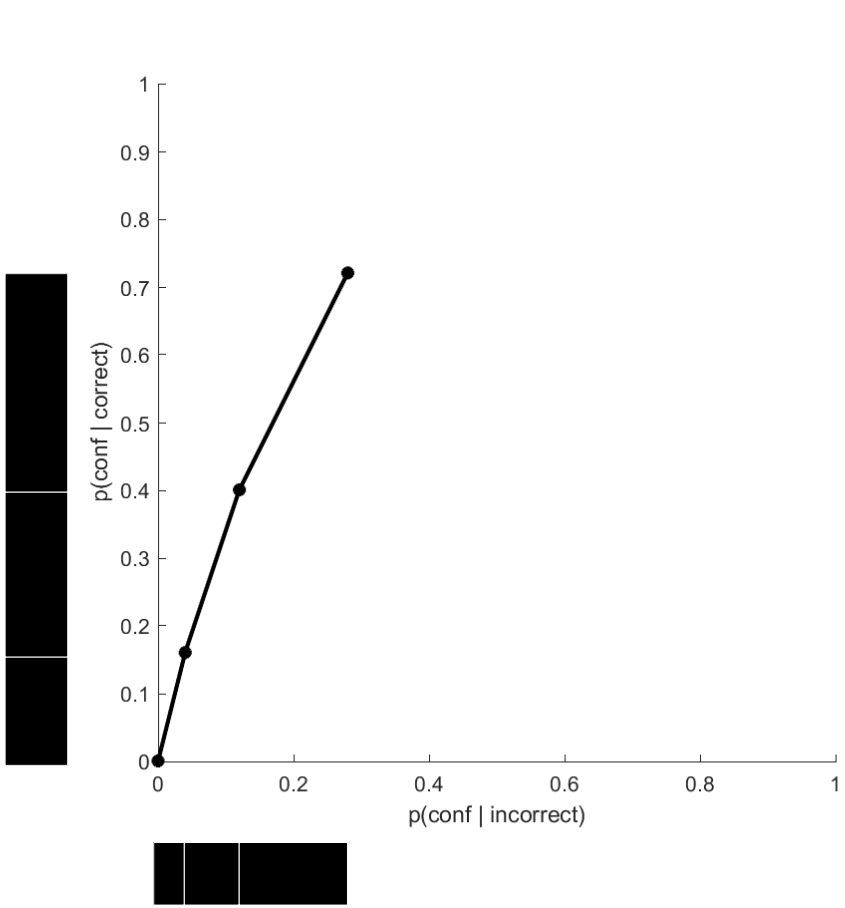

p(conf | correct)
p(conf | incorrect)
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
0.2
0.4
0.6
0.8
1

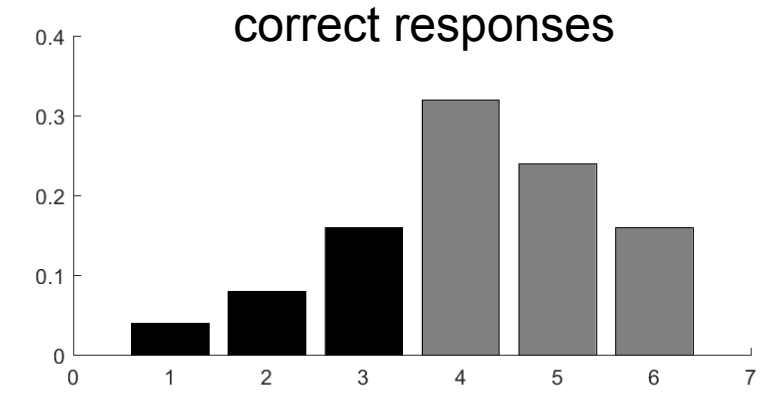

correct responses
incorrect responses
frequency
confidence rating
0
0.1
0.2
0.3
0.4
0
1
2
3
4
5
6
7

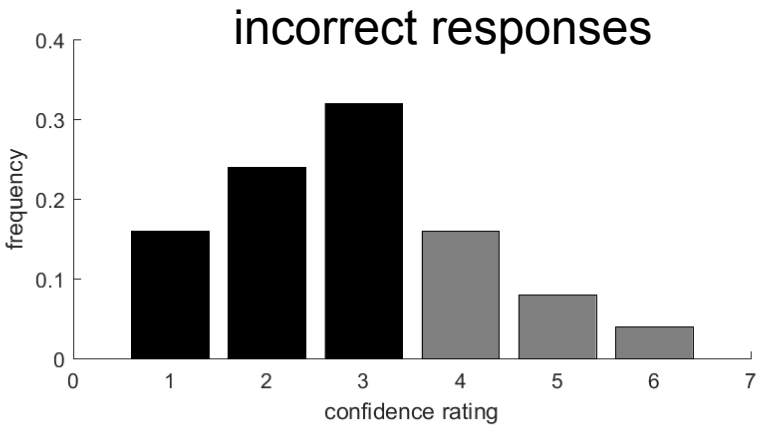
Type 2 ROCs


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

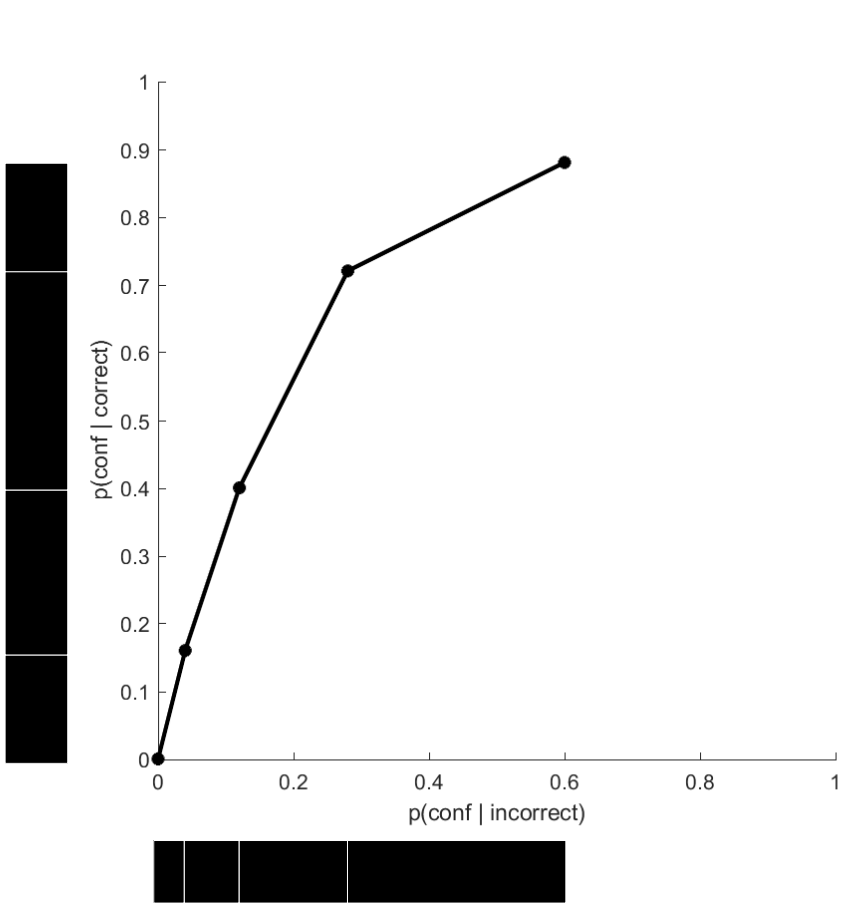
Type 2 ROCs

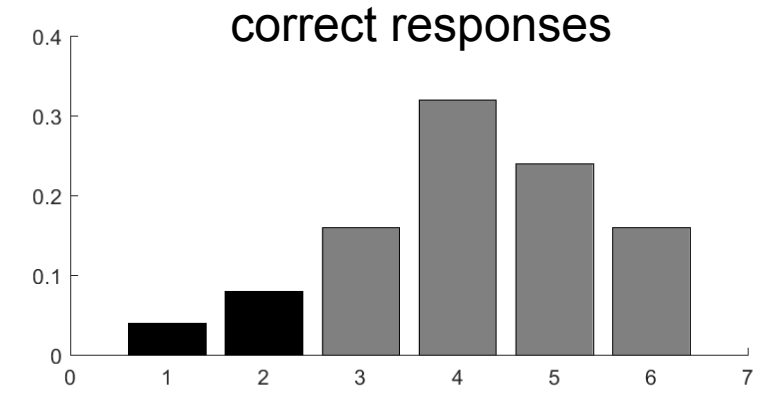

p(conf | correct)
p(conf | incorrect)
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0.2
0.4
0.6
0.8

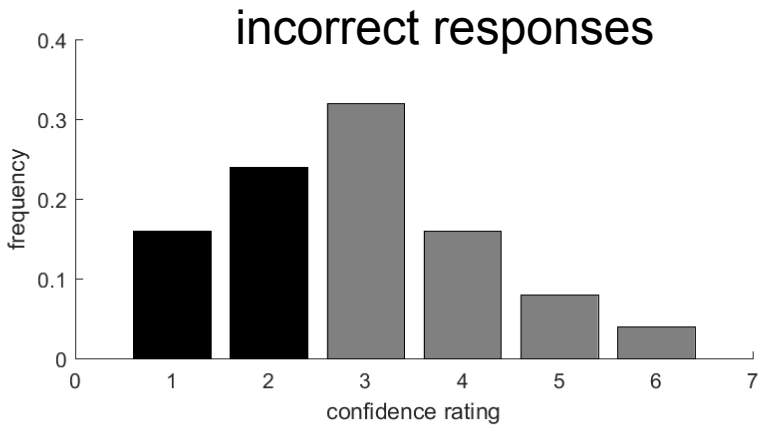

correct responses
0.1
0.2
0.3
0.4
0
1
2
3
4
5
6
7


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

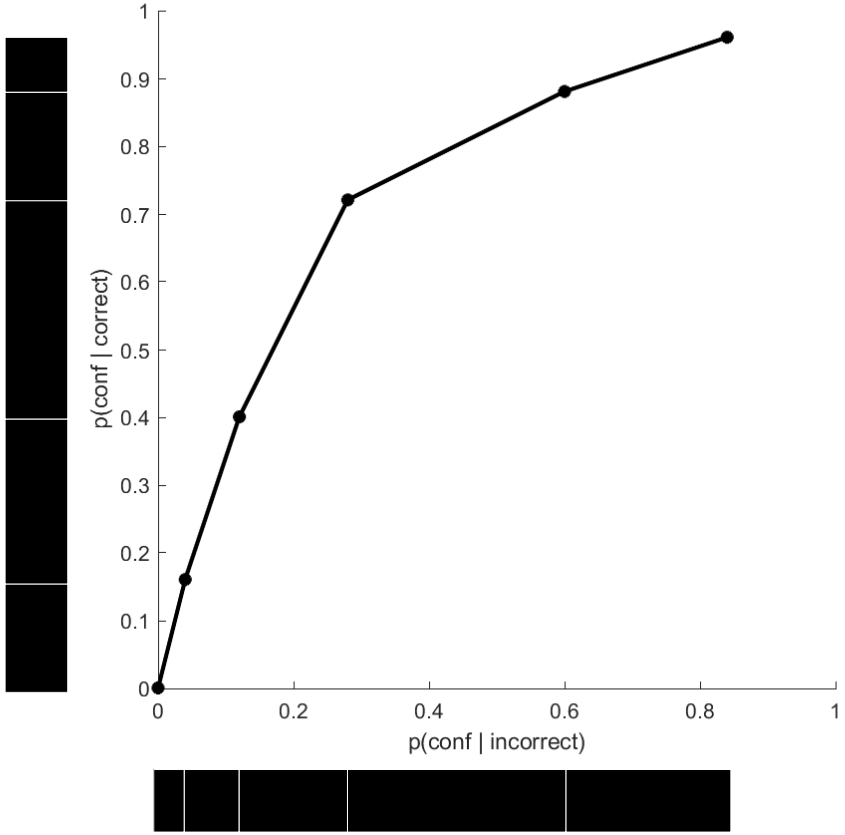
Type 2 ROCs

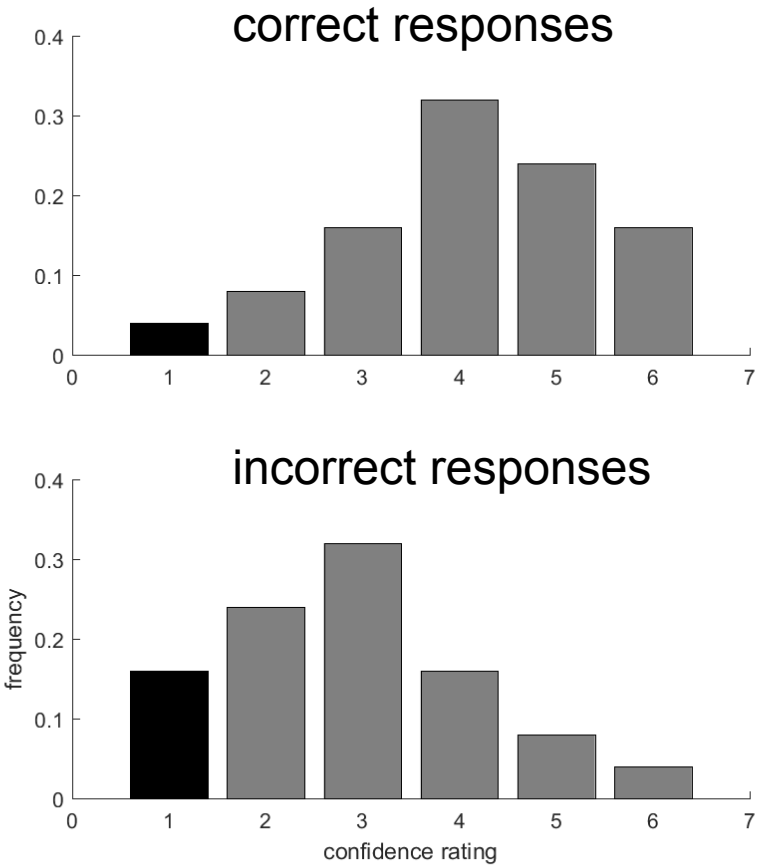

p(conf | correct)
p(conf | incorrect)
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
0.2
0.4
0.6
0.8
1


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

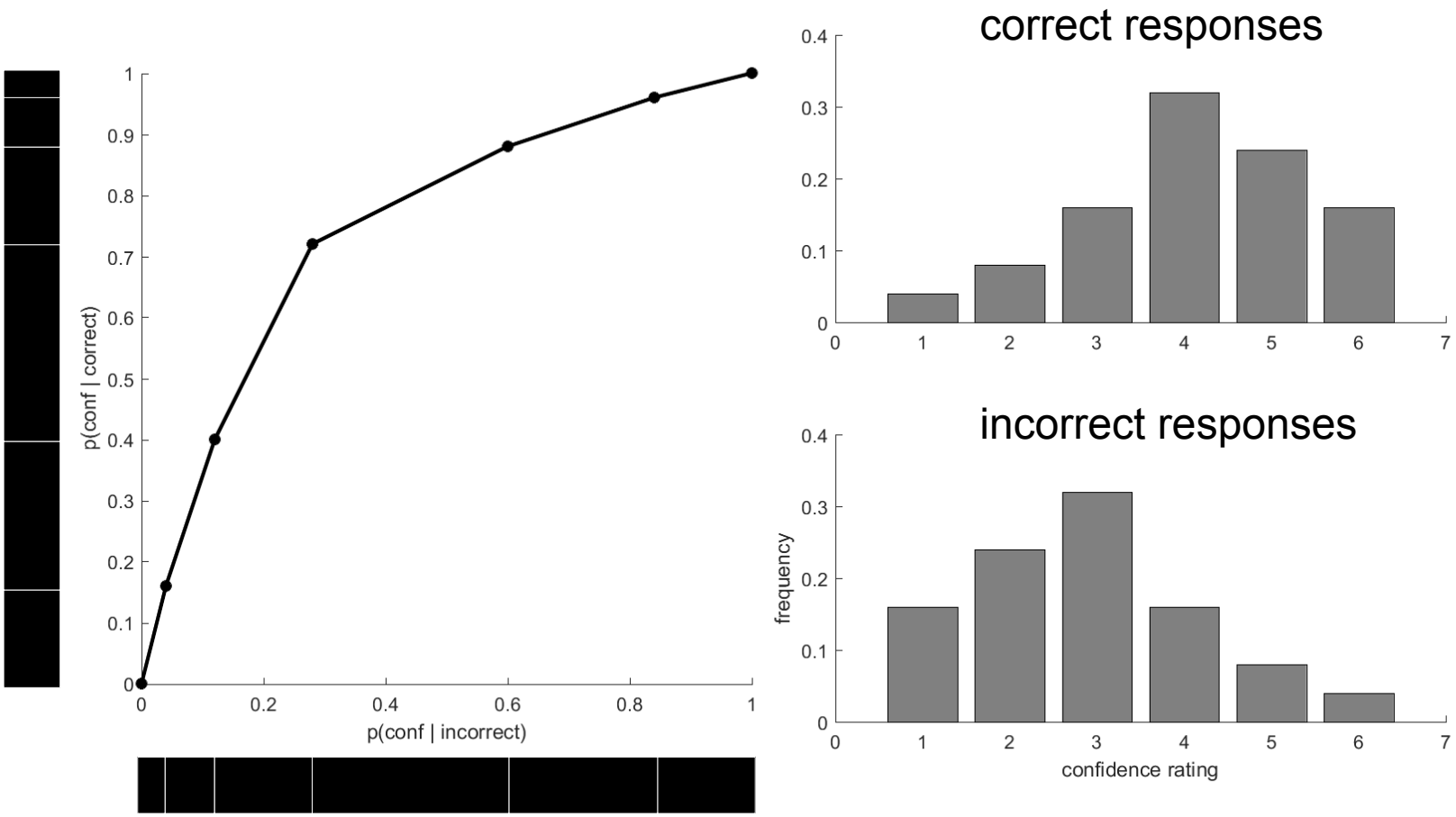

incorrect responses
frequency
0.4
0.3
0.2
0.1
0
0
1
2
3
4
5
6
7
confidence rating

Type 2 ROCs

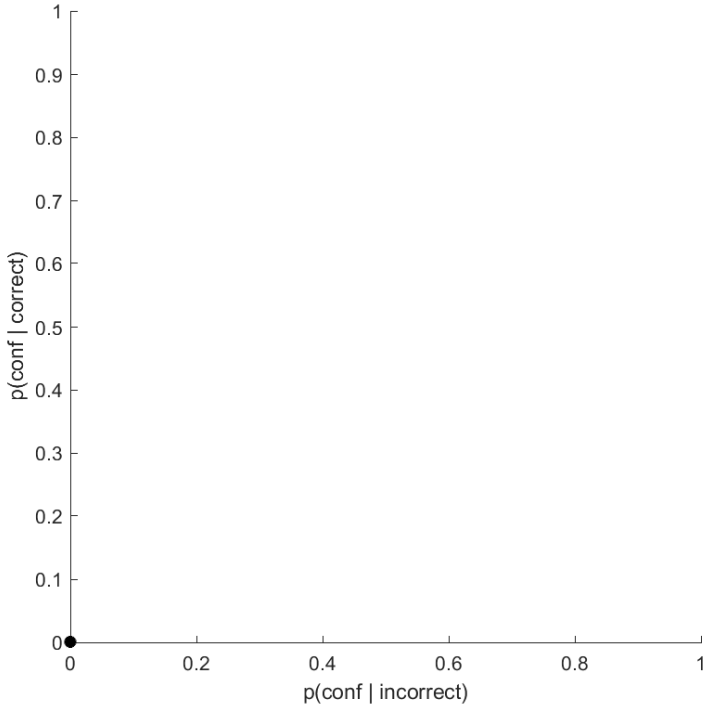

p(conf | correct)
p(conf | incorrect)
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
0.2
0.4
0.6
0.8
1


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

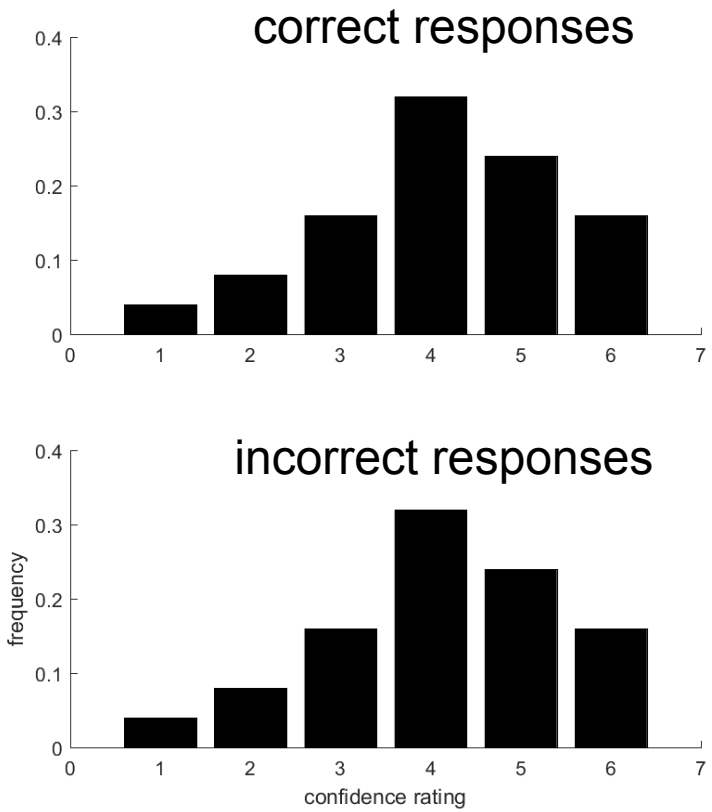
Type 2 ROCs


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

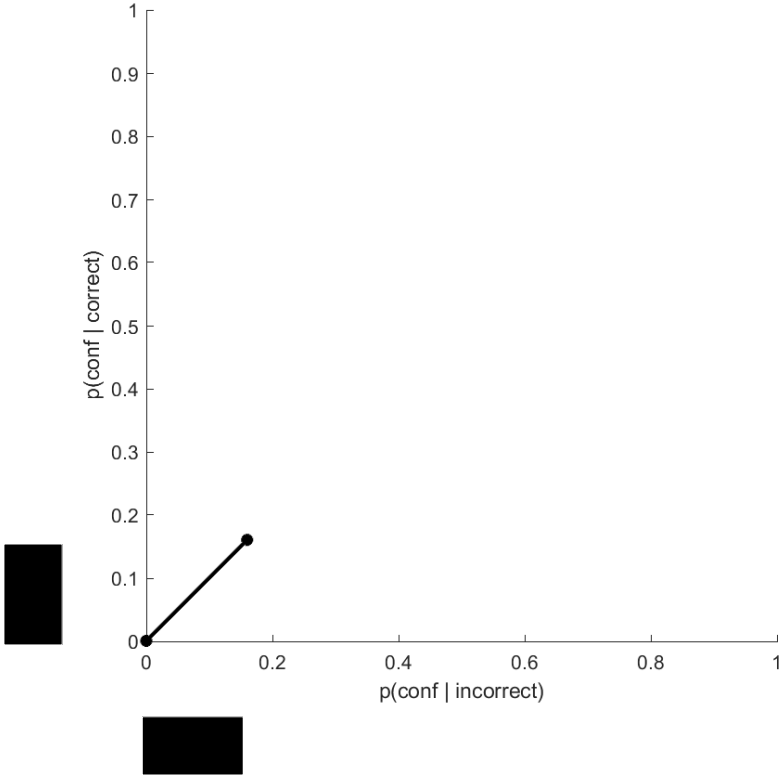
Type 2 ROCs



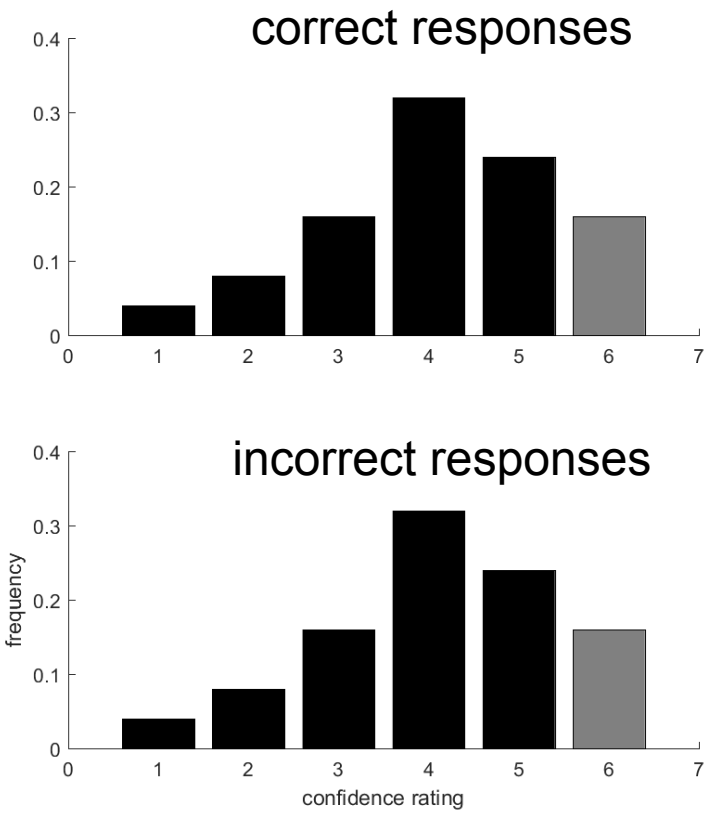
No metacognition:



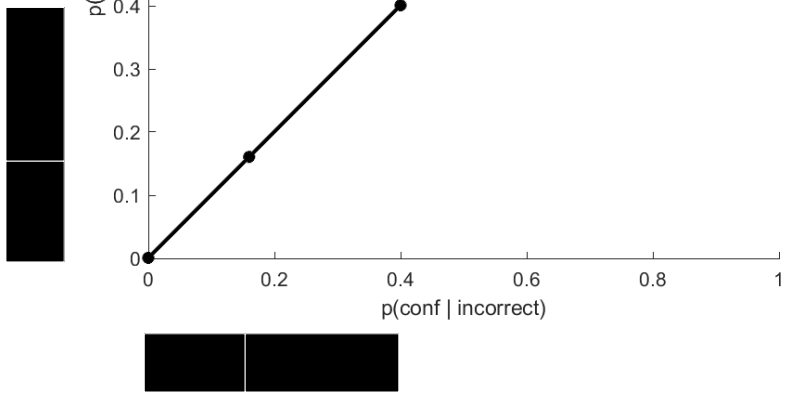
Type 2 ROCs



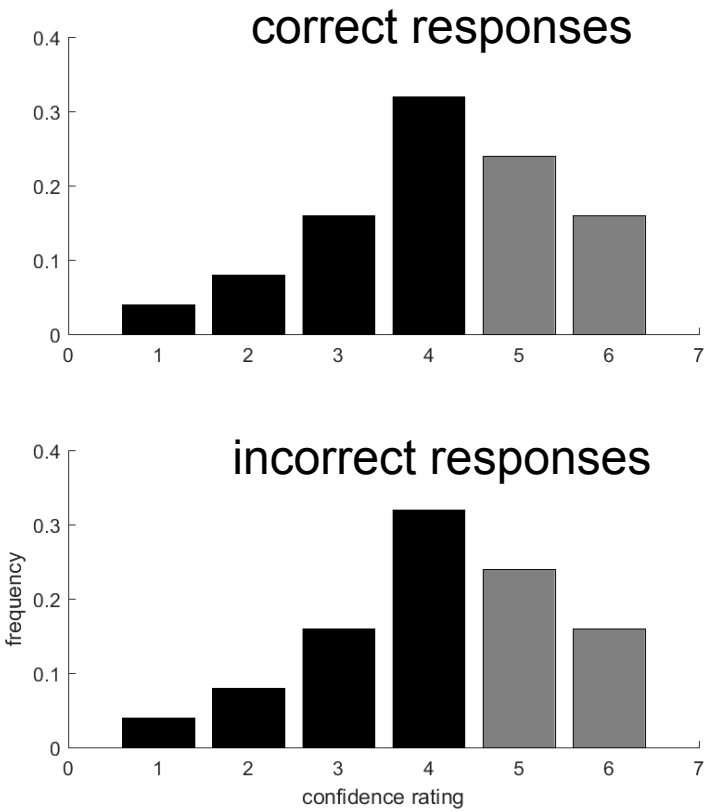
No metacognition:



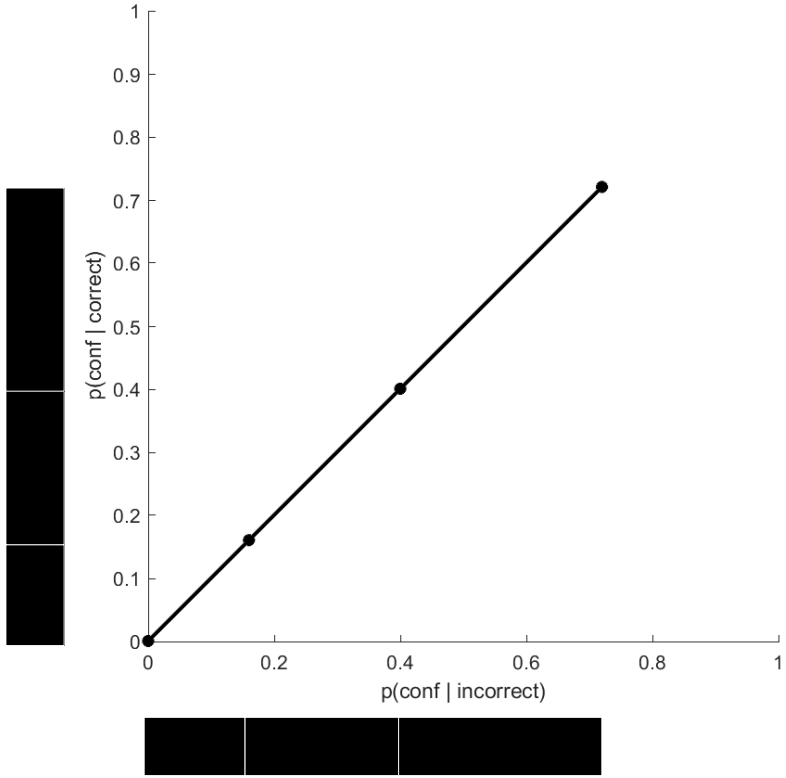
Type 2 ROCs



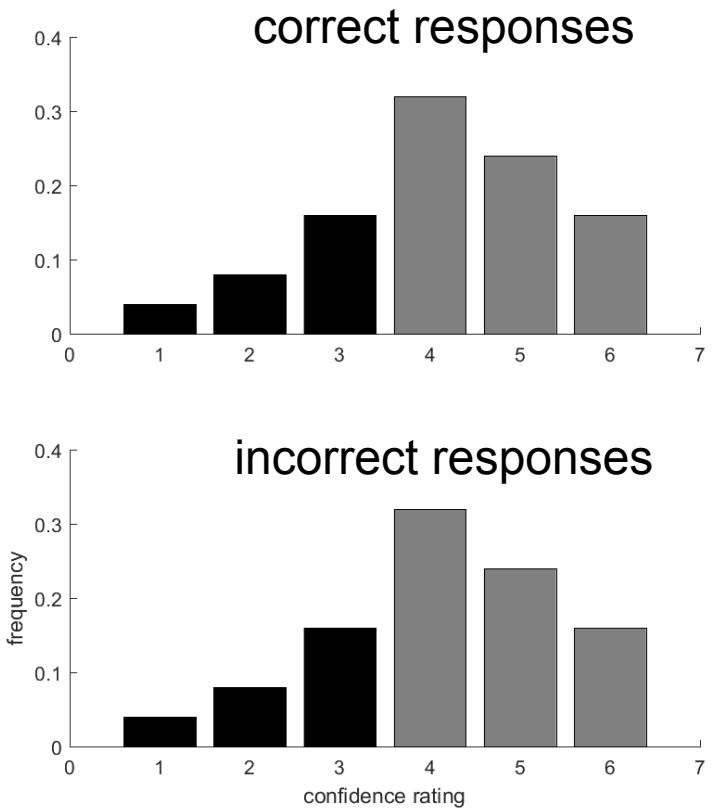
No metacognition:



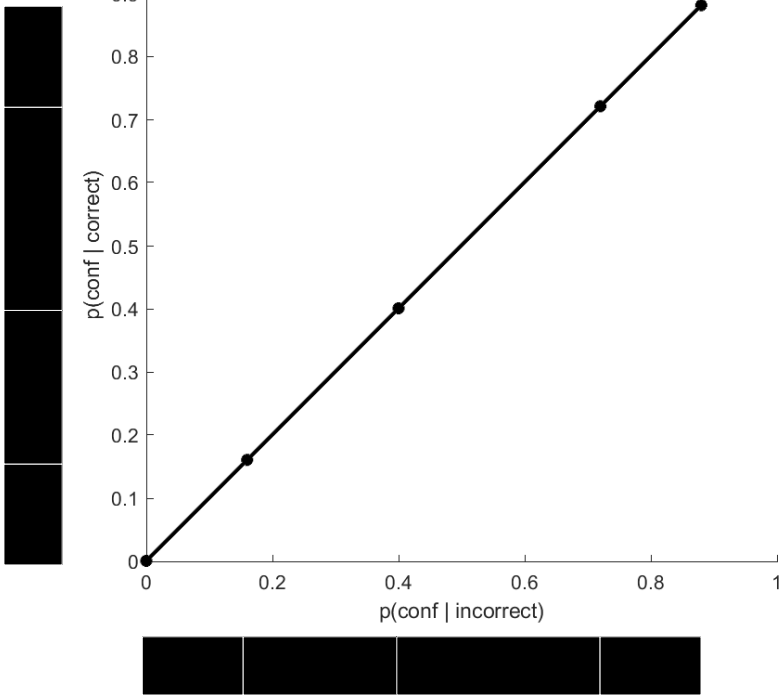
Type 2 ROCs



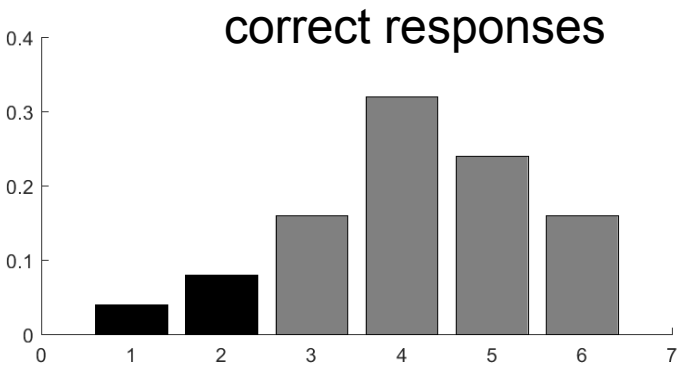
No metacognition:

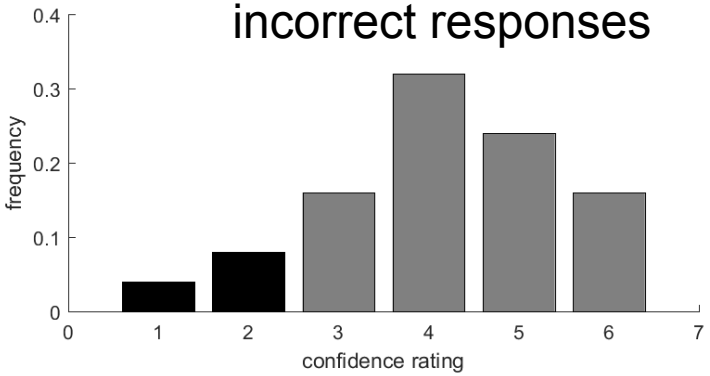


Type 2 ROCs

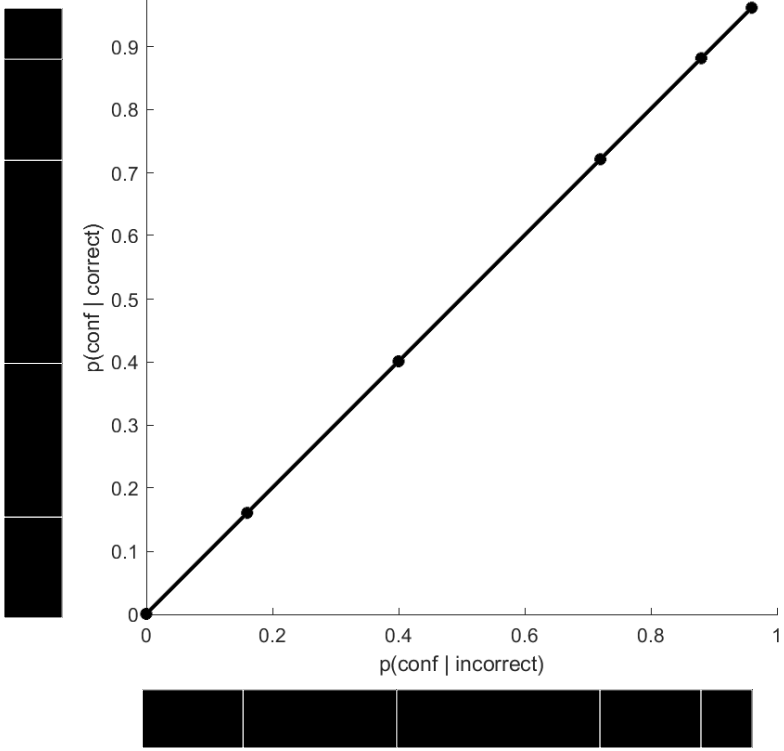


No metacognition:

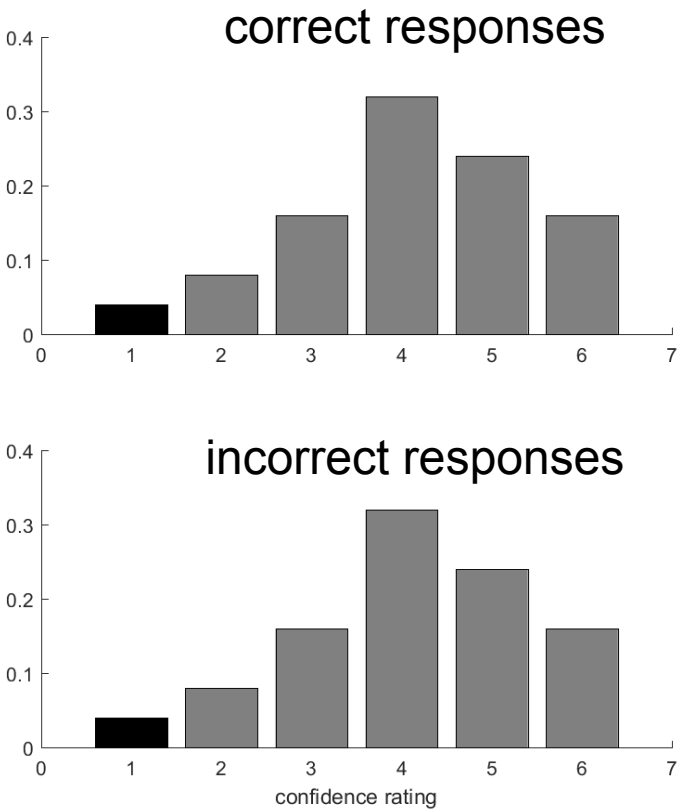




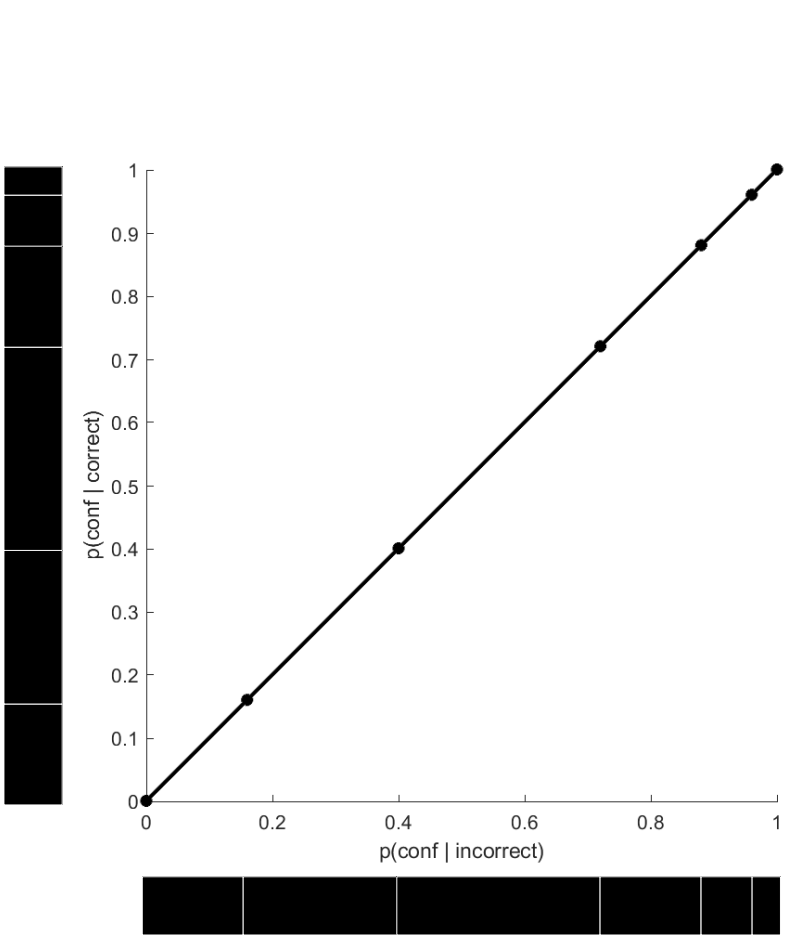
Type 2 ROCs



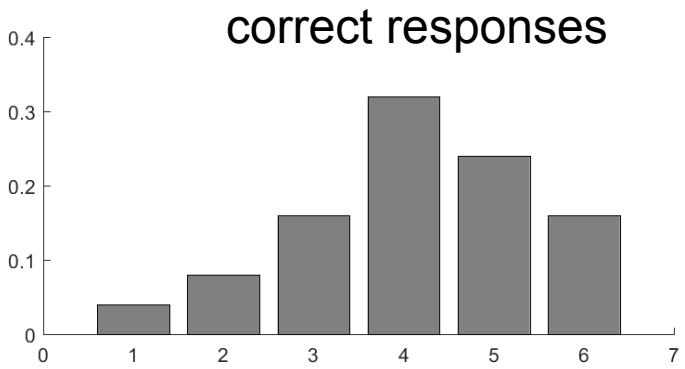
No metacognition:

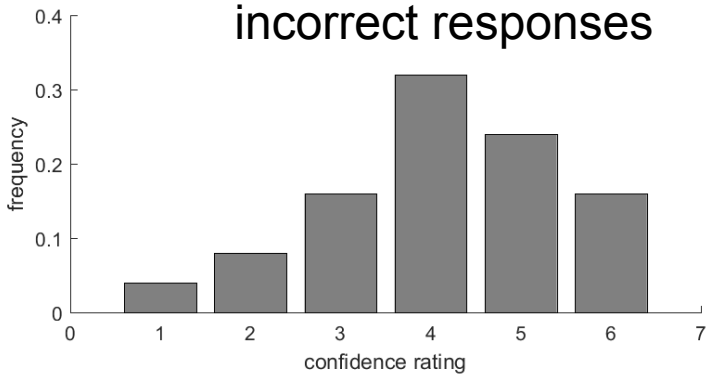


Type 2 ROCs



No metacognition:





Type 2 ROCs

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

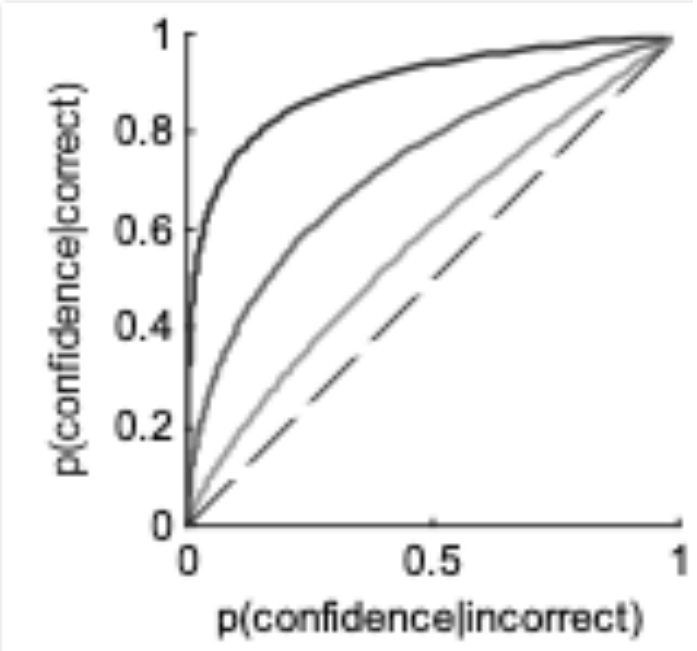
Simulated type 2 ROC

Type 1 performance

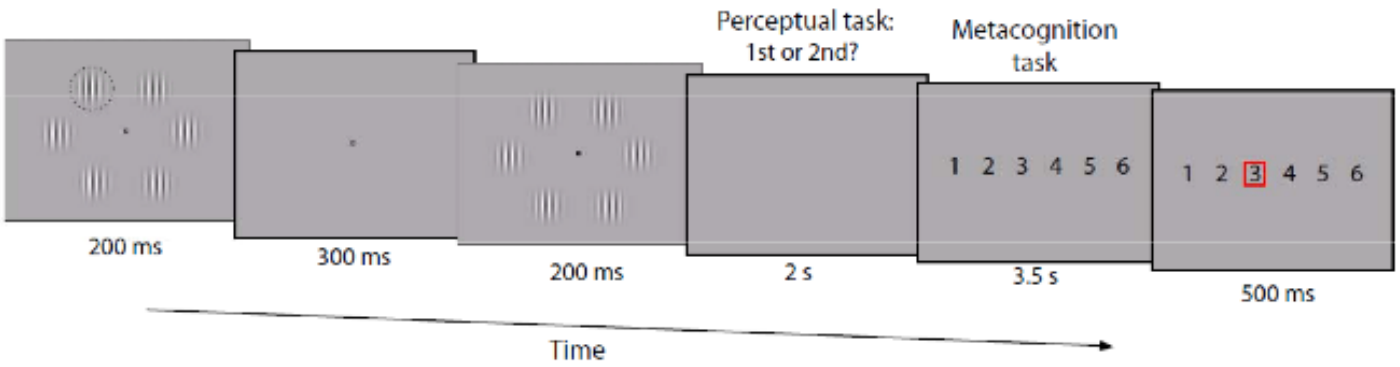
d' = 0.5

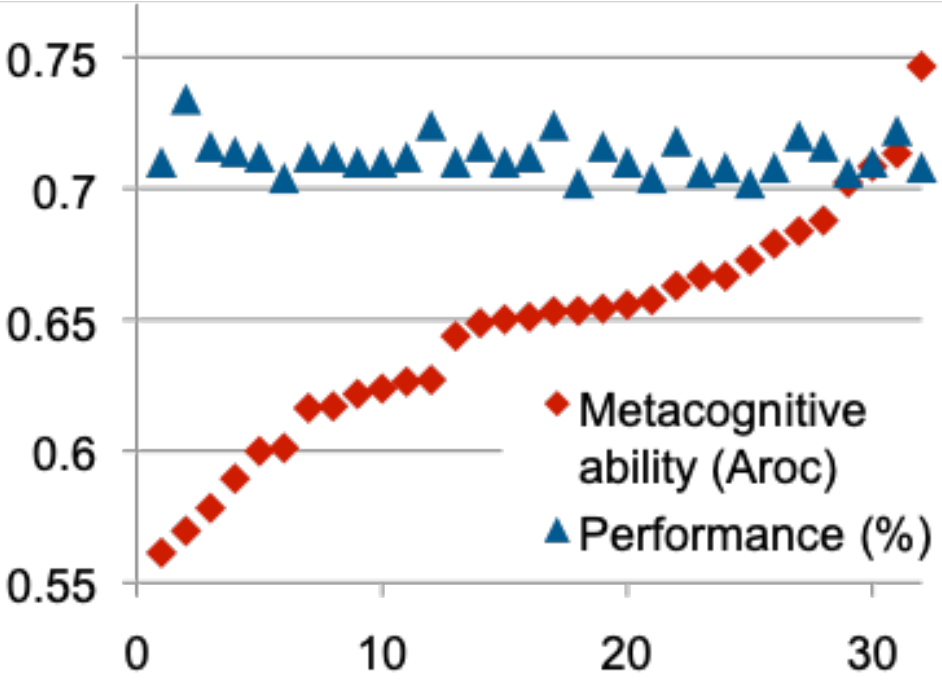
d' = 1.5

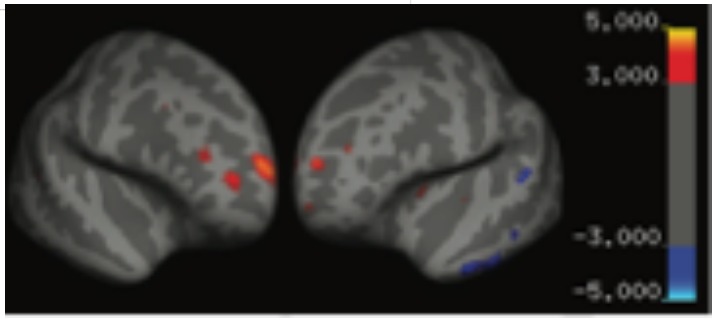
d' = 3.0

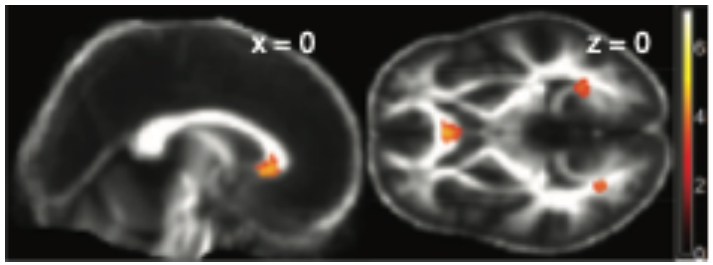


Empirical example







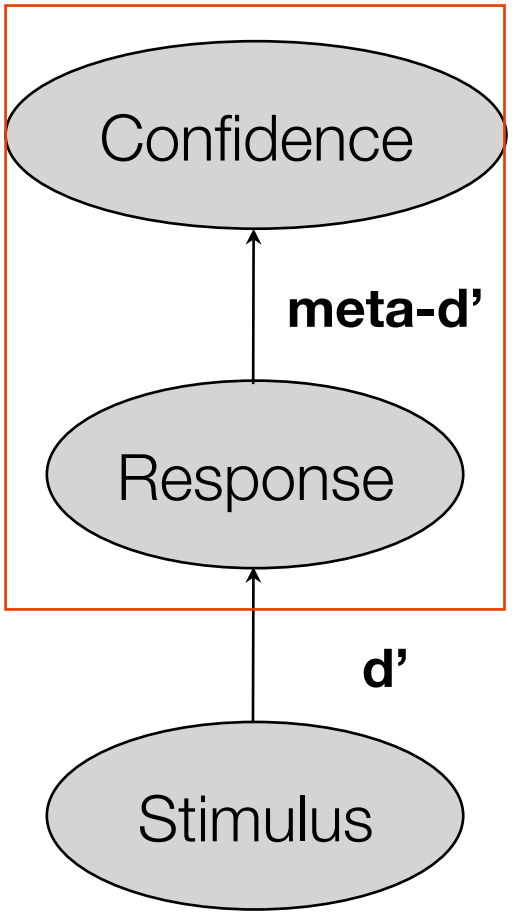


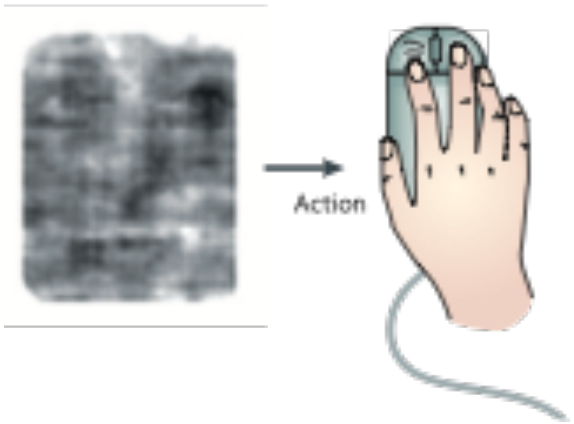
Fleming et al. (2010) *Science*

Quantifying metacognition (3) - meta-d’

Metacognitive sensitivity

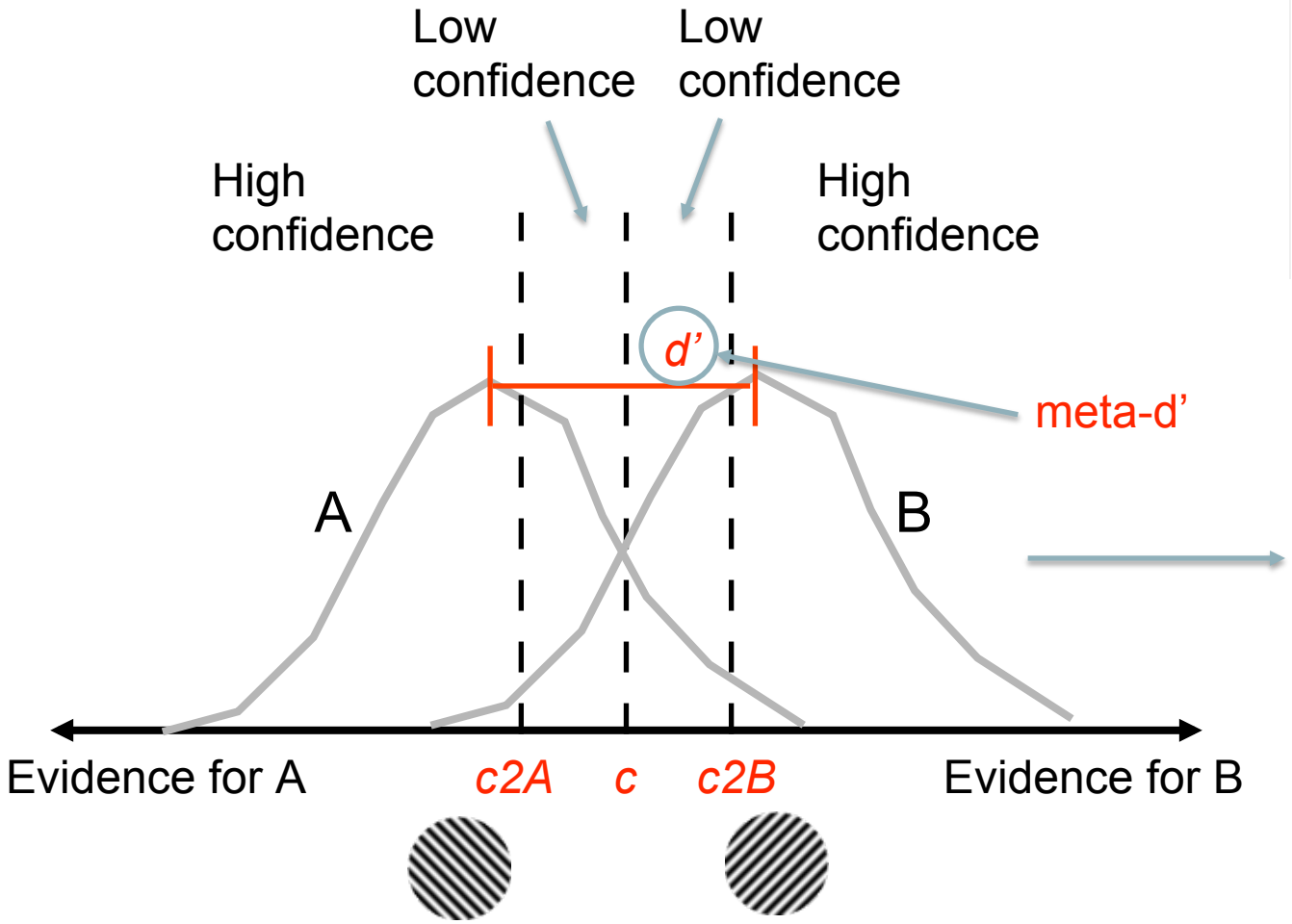
First-order sensitivity



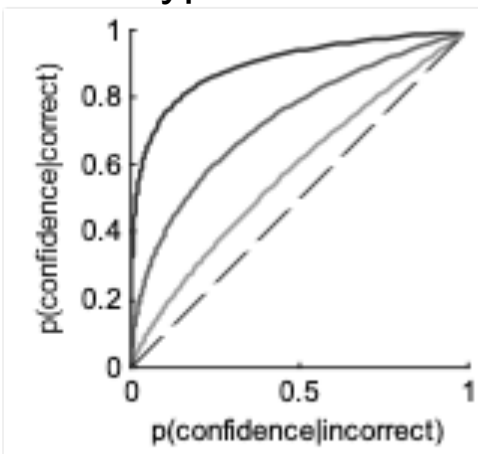


Maniscalco & Lau (2012) *Consciousness & Cognition*; Fleming (2017) *Neurosci of Consciousness*

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*

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.

✓ 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 – a desirable feature.

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

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

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, evaluation of test-retest reliability...

... have been applied in computational psychiatry

Metacognition and psychopathology

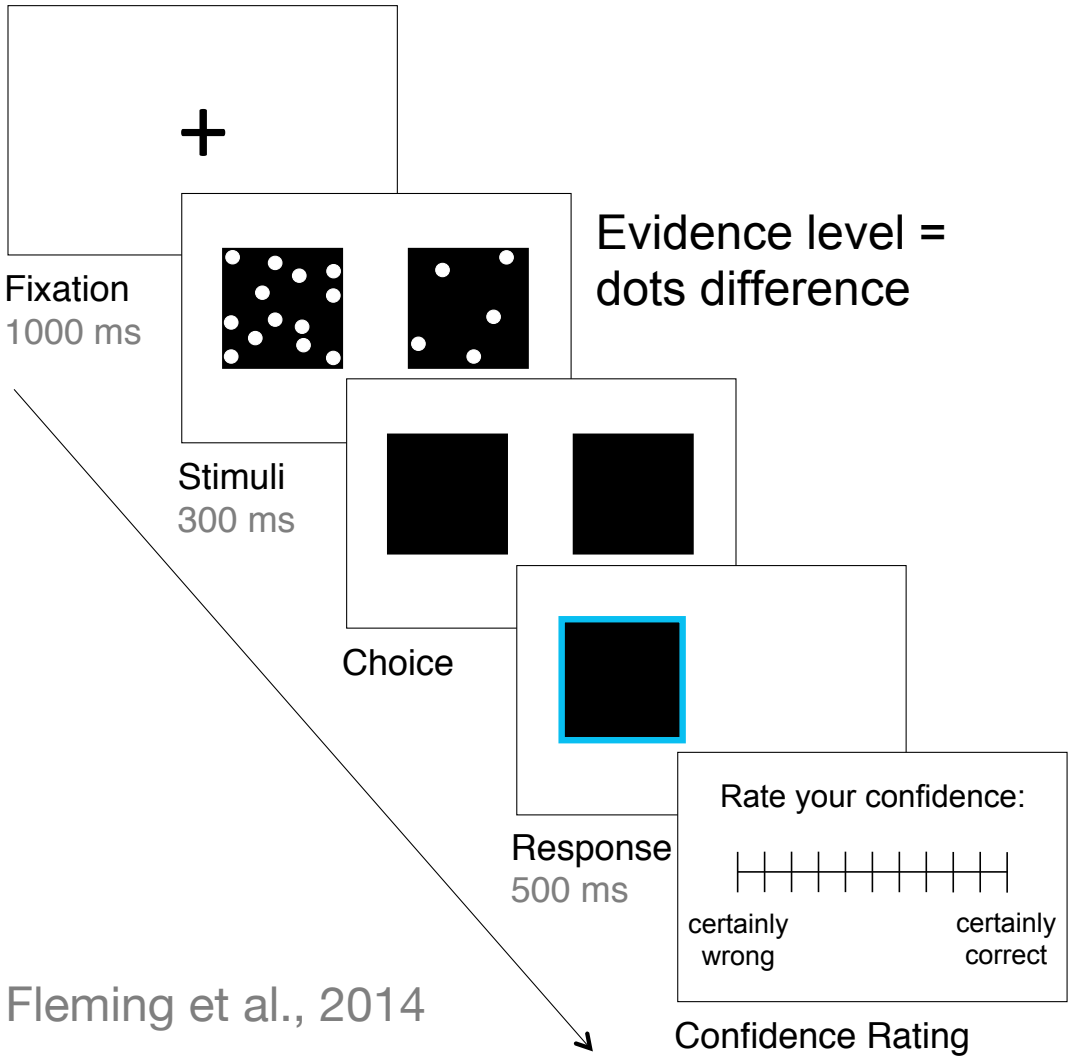
A ‘general population’ sample
Total N=995 participants

Perceptual decision-making task

Self-reported symptom questionnaires

Rouault*, Seow*, Gillan and Fleming (2018) *Biological Psychiatry*

Perceptual decision-making task



Fleming et al., 2014

QUANTIFY

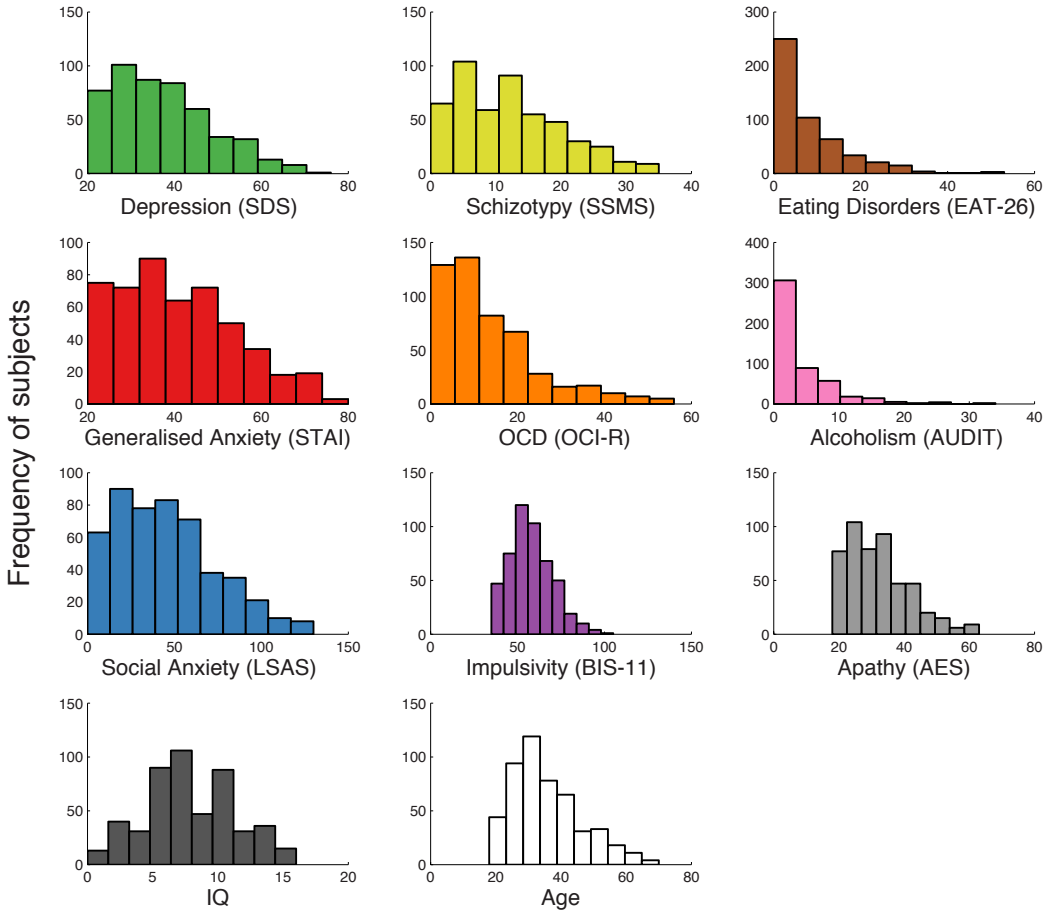
- **Decision process**

Accuracy
(Drift-diffusion model)

- **Metacognition**

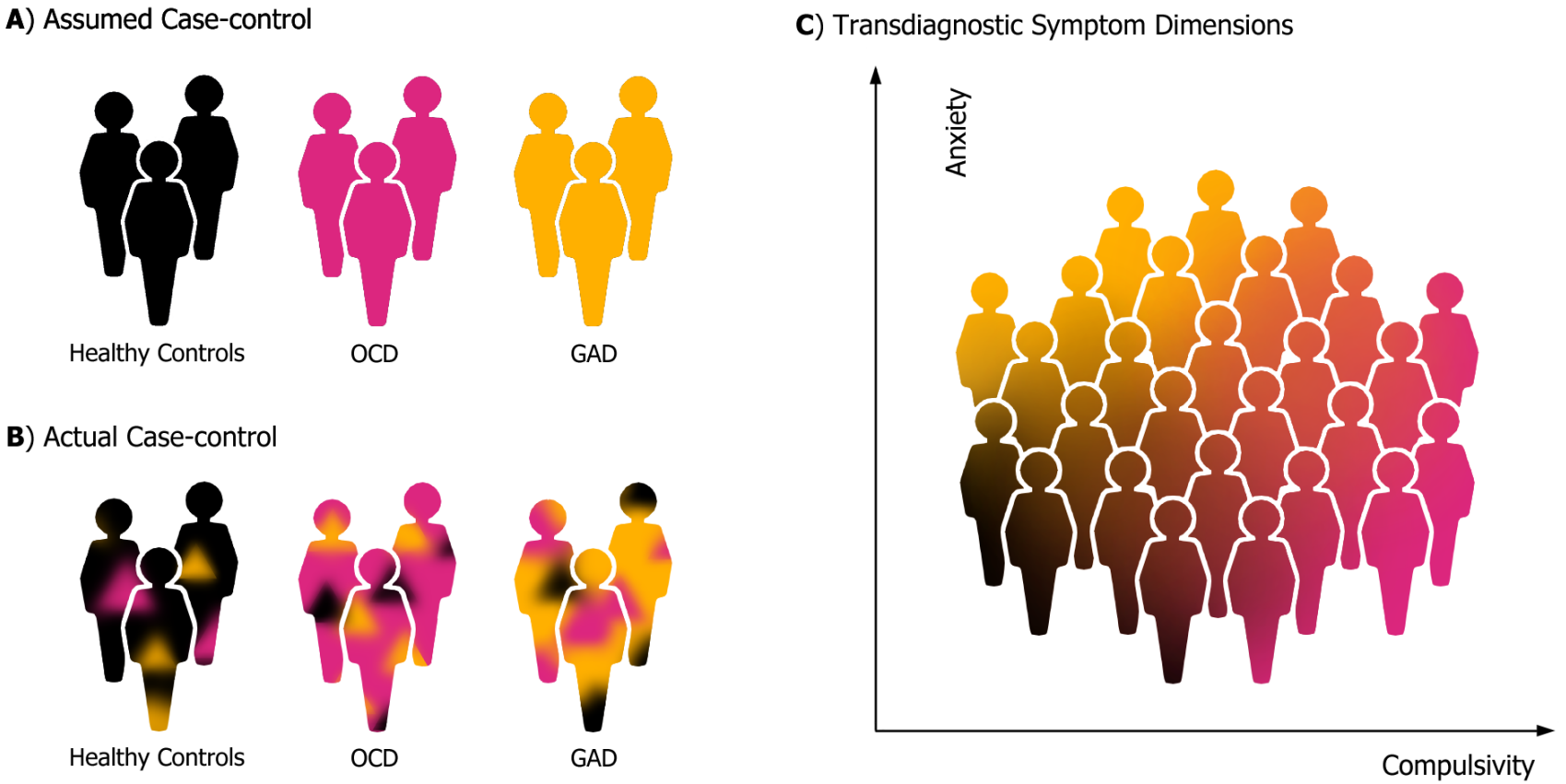
⇒ **Confidence level**
⇒ **Metacognitive efficiency**

Self-reported psychiatric symptoms



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

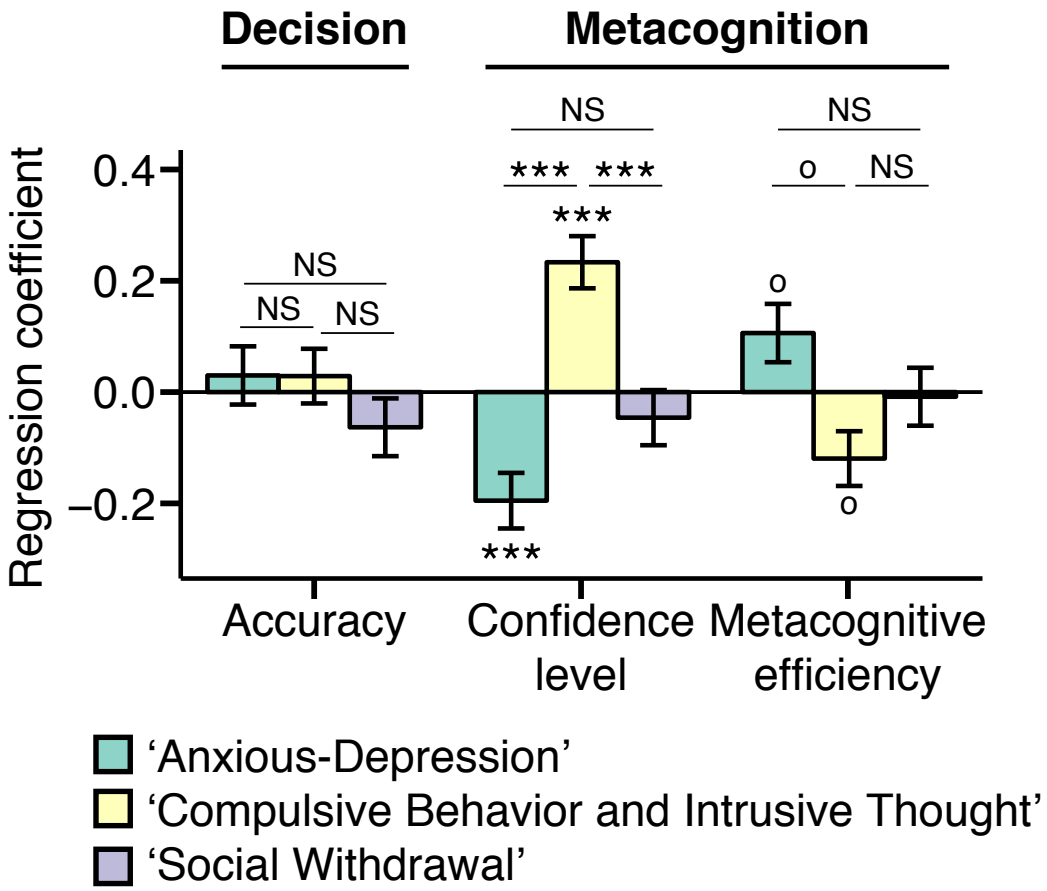
A transdiagnostic approach



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

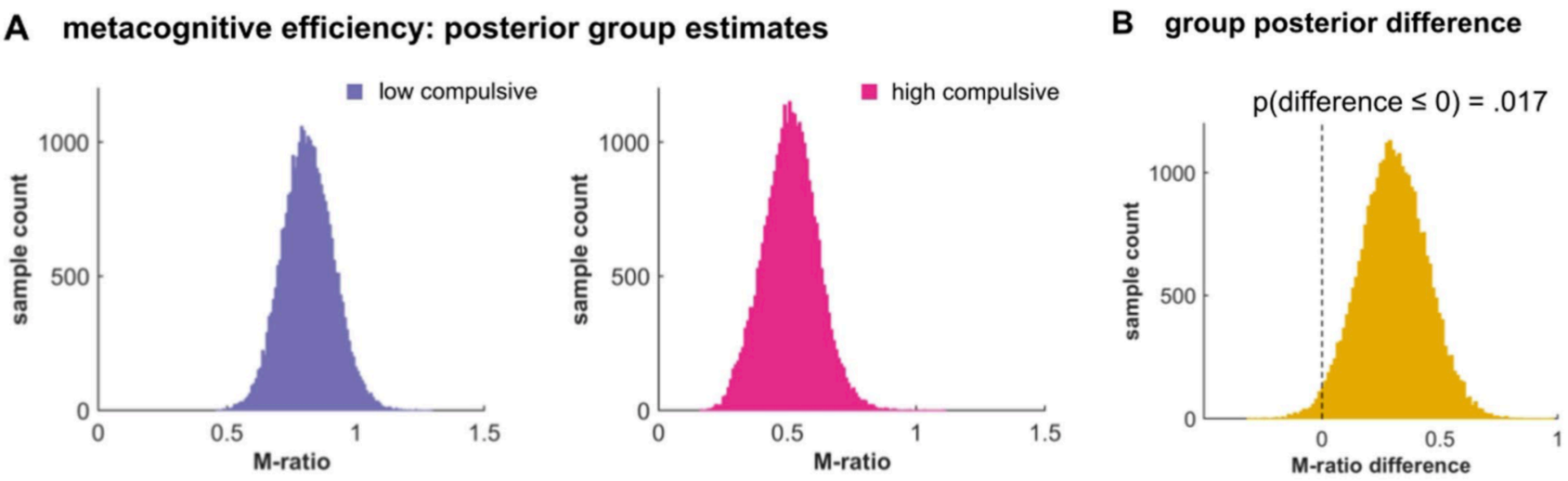
Confidence ↓
Metacognitive efficiency ↑

Compulsive/Intrusive

Confidence ↑
Metacognitive efficiency ↓

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

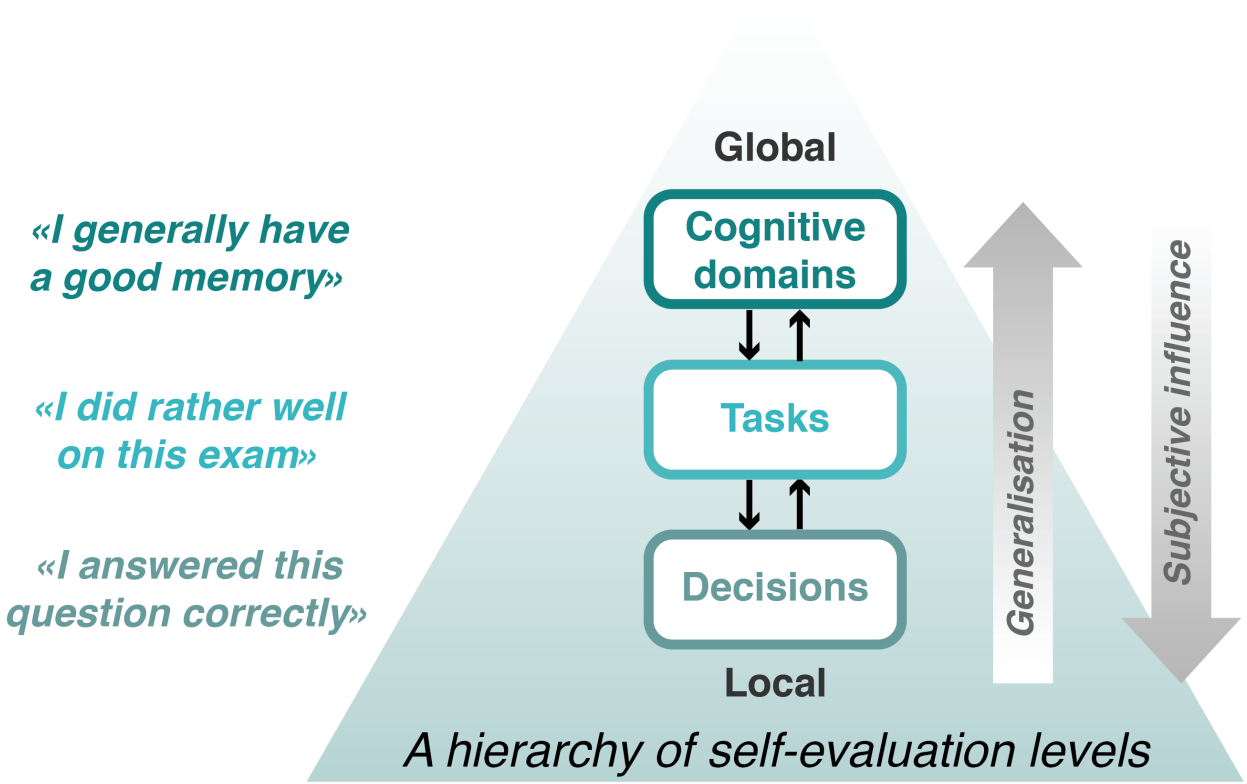
Metacognition and computational psychiatry



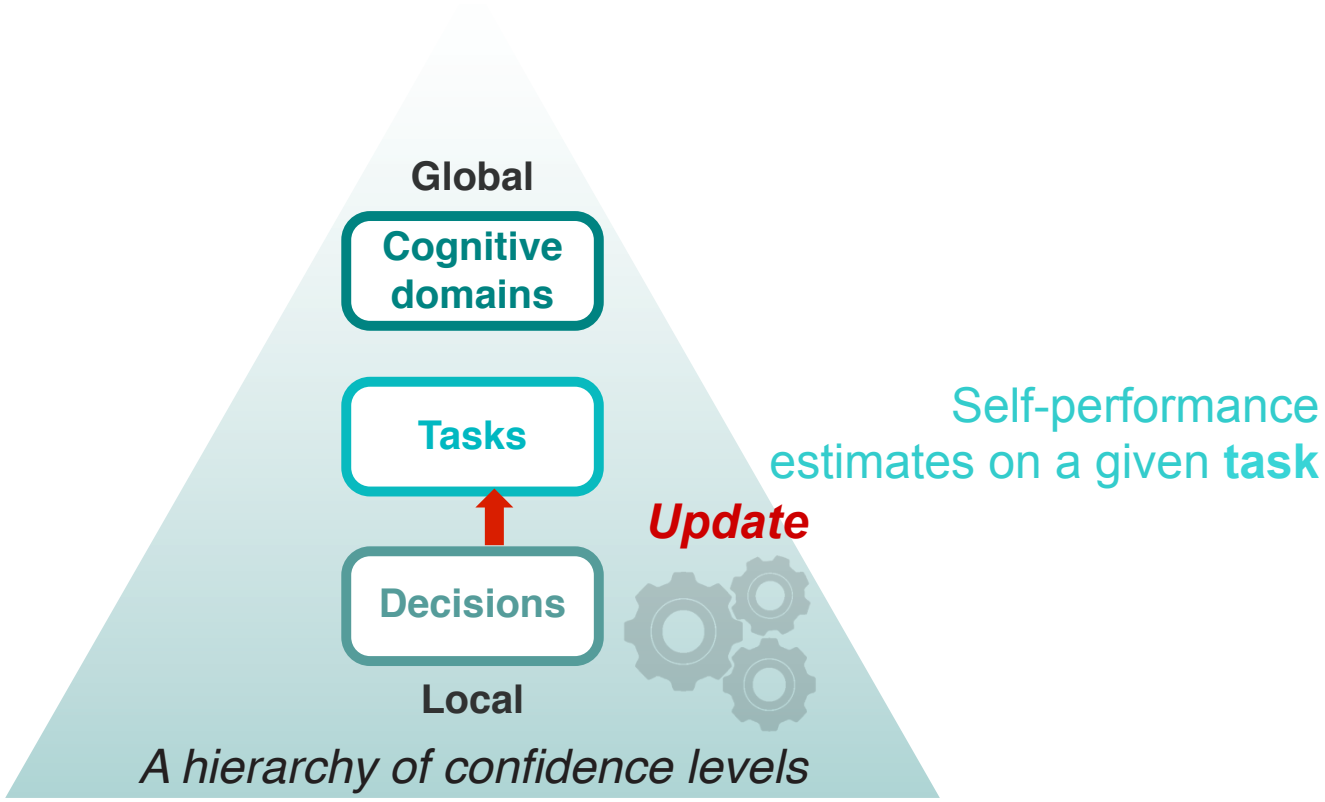
Reduced perceptual metacognitive efficiency in individuals with high compulsivity (N=20 per group)

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

Metacognition operates across a hierarchy of levels

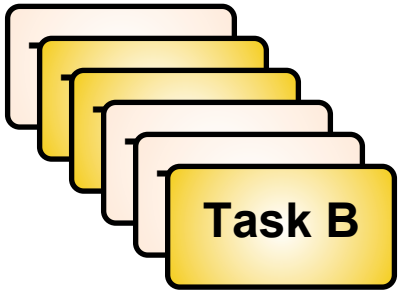


Seow*, Rouault*, Gillan & Fleming (2022) *Biological Psychiatry*


Global
Cognitive domains
Tasks
Update
Decisions
Local
A hierarchy of confidence levels
Self-performance estimates on a given task

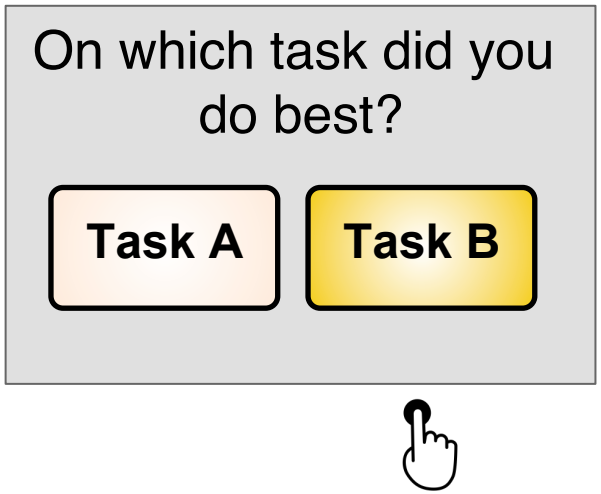
Confidence at the level of task (“global”)

Mini-learning blocks



Global confidence
Measures:

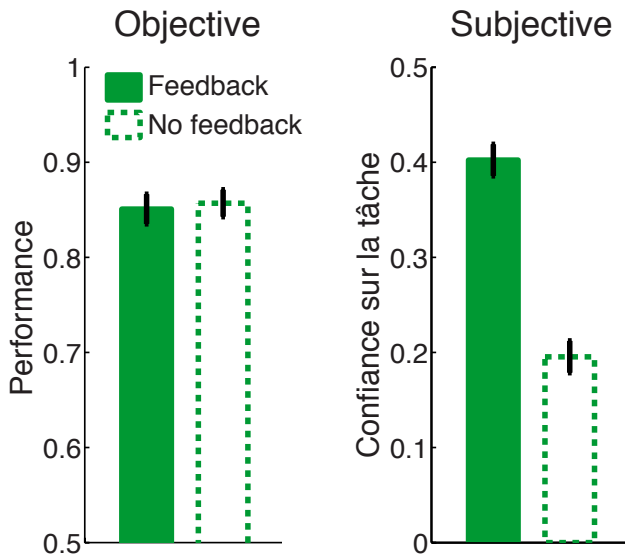
- Choice between tasks (retrospective or prospective)
- Task ratings



Rouault, Dayan & Fleming, 2019 *Nat Commun;* Rouault & Fleming, 2020 *PNAS*

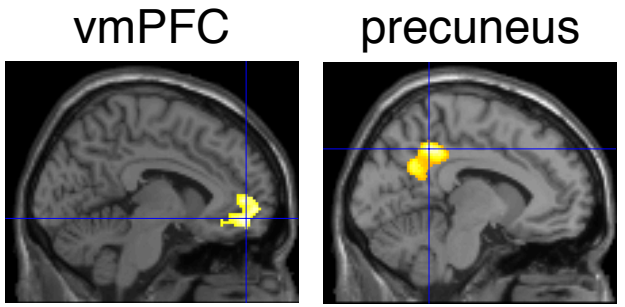
Confidence at the level of task (“global”)

Internal metacognitive evaluation and **external** feedback influence the formation of self-confidence on a given task

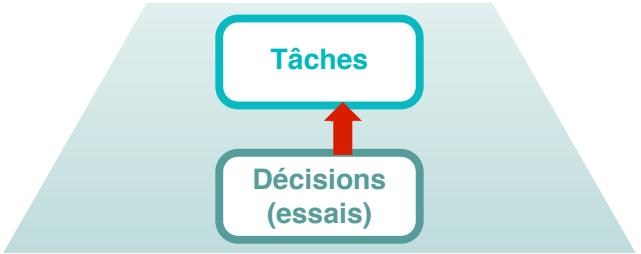


Participants underestimate their abilities in the absence of feedback, despite objective performance remaining unchanged

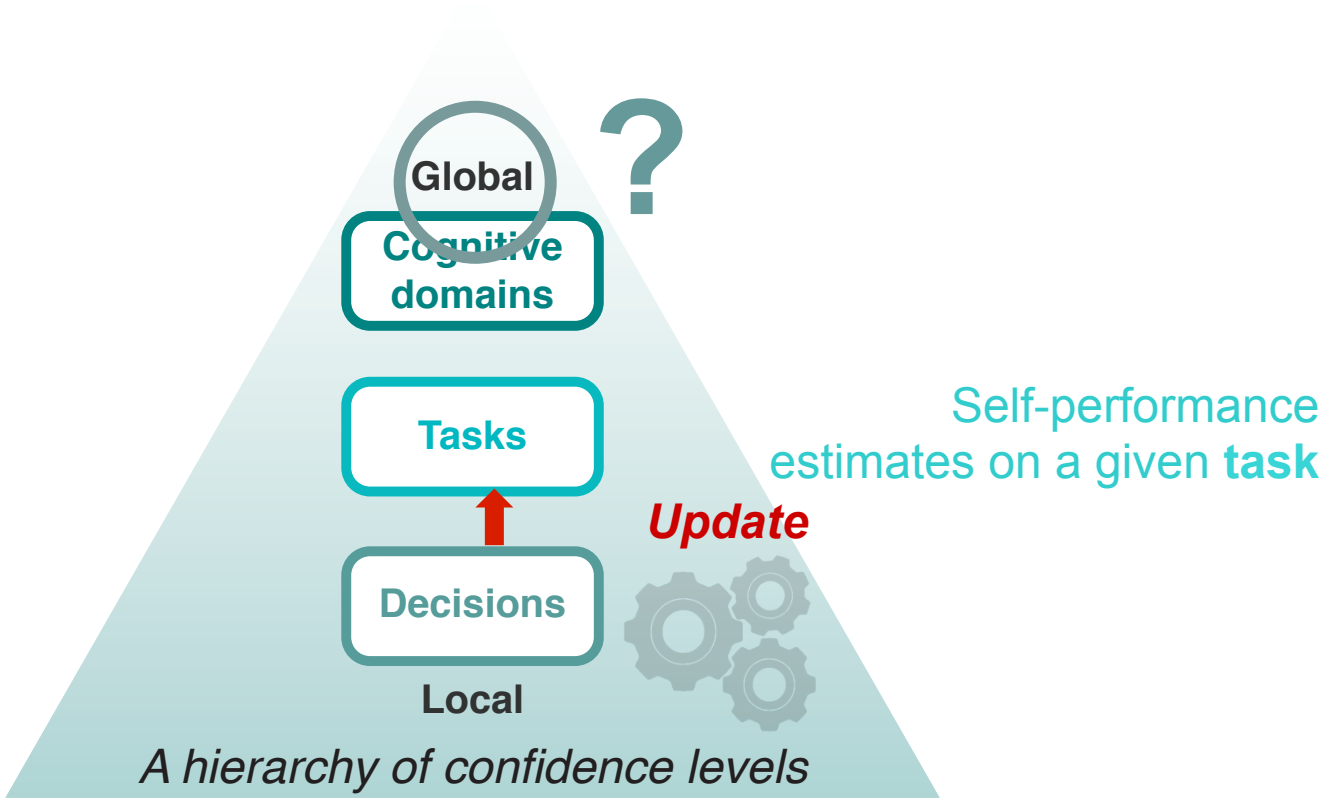
Rouault, Dayan & Fleming, 2019 *Nat Commun*



Confidence is signalled across various hierarchical levels across the human brain



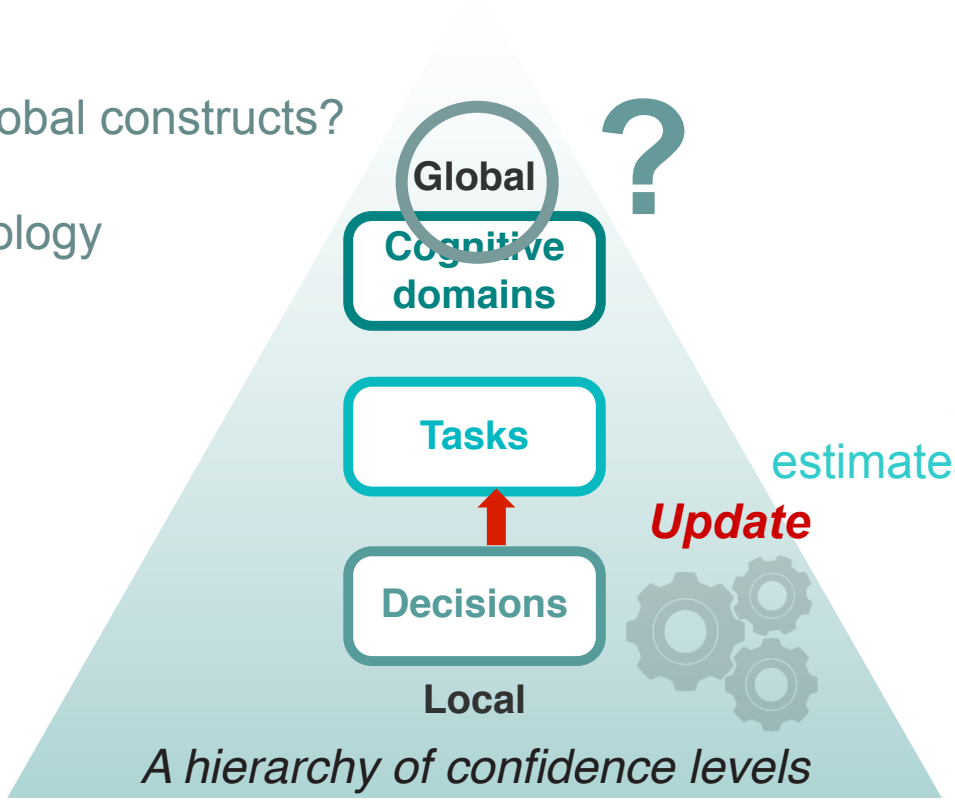
Rouault & Fleming, 2020 *PNAS*


Global
?
Cognitive domains
Tasks
Update
Decisions
Local
A hierarchy of confidence levels
Self-performance estimates on a given **task**

Relevance to global constructs?

A. Self-esteem

B. Psychopathology



Self-performance estimates on a given **task**

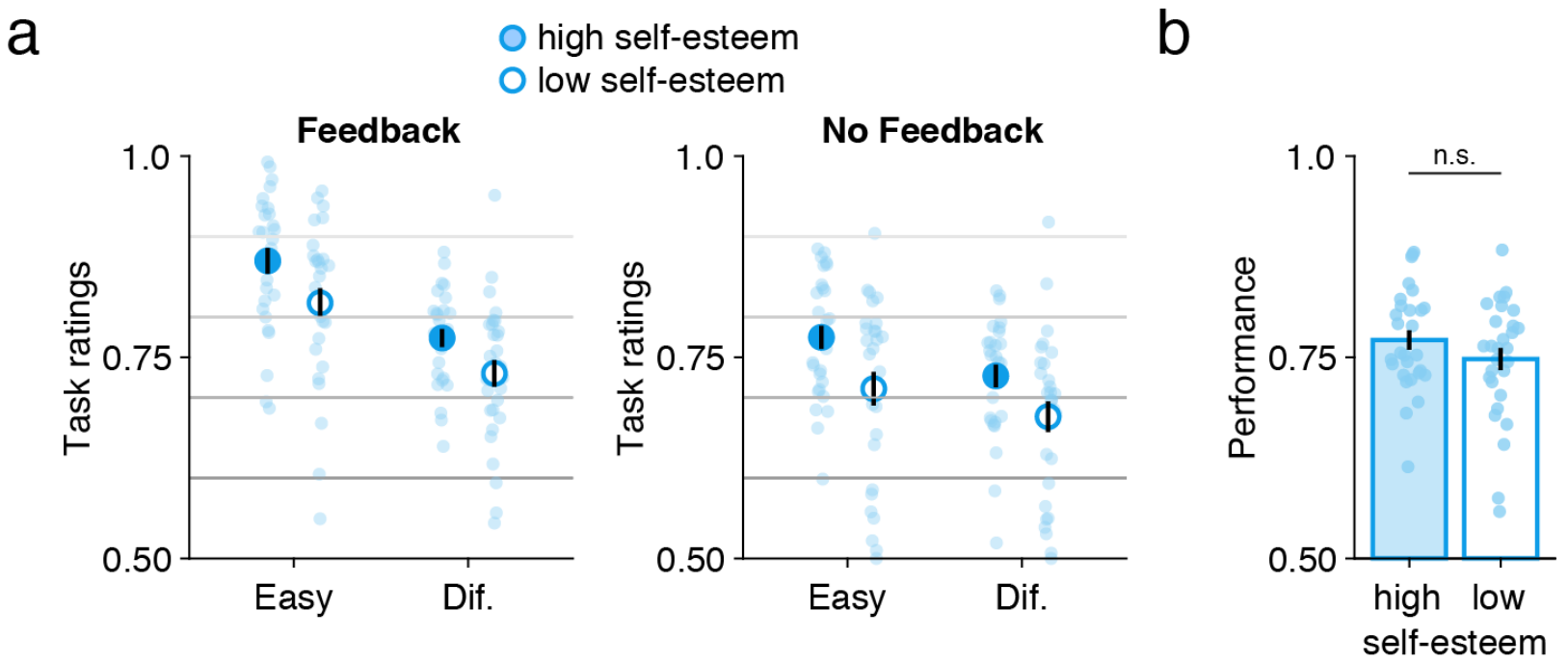
Self-esteem and the formation of global self-performance estimates

» High self-esteem is a cornerstone of mental health

Bandura, 1977; Orth et al., 2008; Elliott et al., 1996

» **Are self-performance estimates on task relevant for understanding global constructs such as self-esteem?**

Self-esteem and the formation of global self-performance estimates



ANOVA on **task ratings**:
Main effect of self-esteem

ANOVA on **performance**:
No main effect of self-esteem

Rouault*, Will*, Fleming & Dolan (2022) *Translational Psychiatry*

Metacognition and mental health symptoms

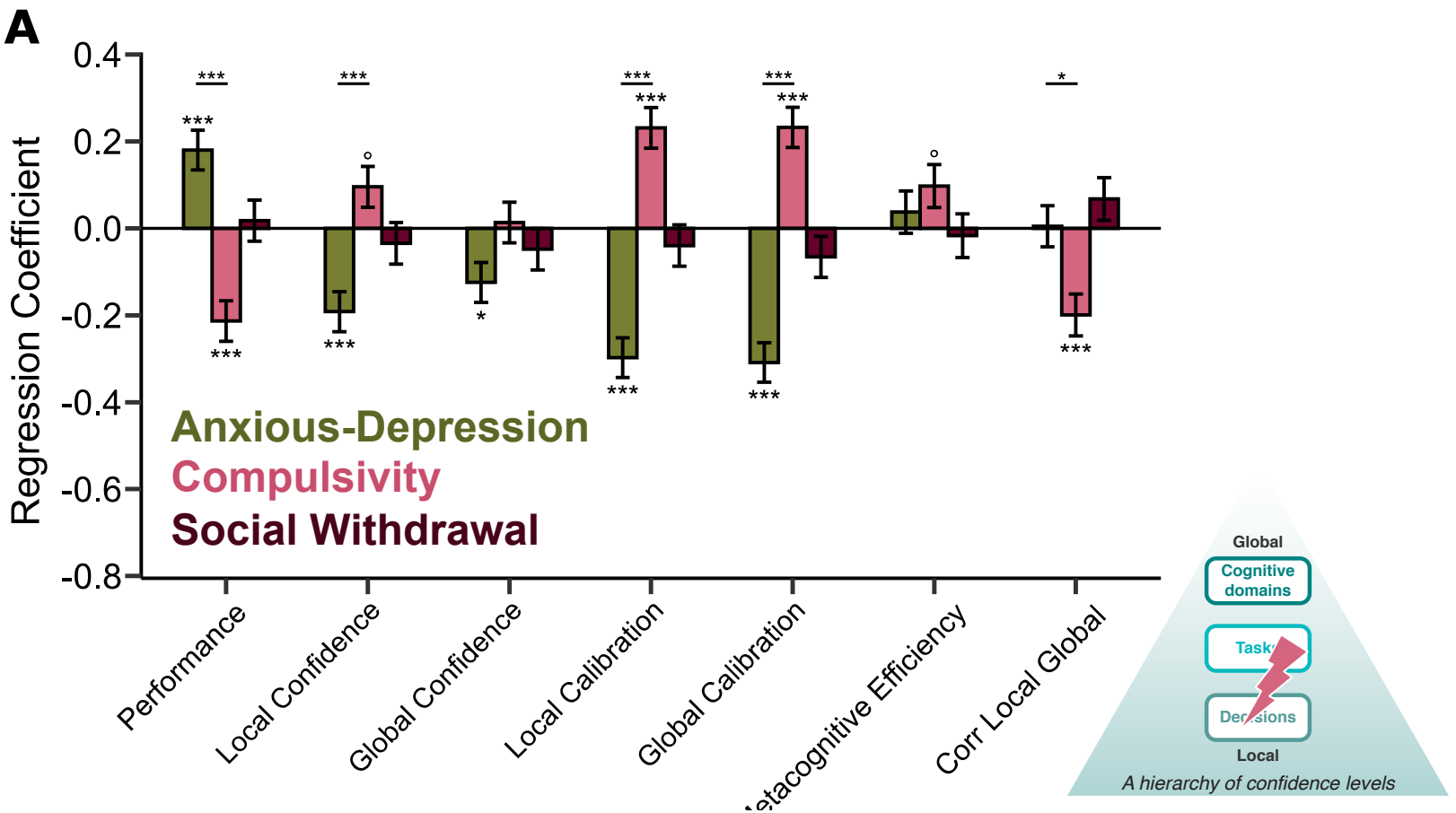
Prolific N=489 participants

Local-Global metacognition task

Self-reported symptom questionnaires

Hoven, Denys, Luigjes, Rouault & van Holst (2023) *Nature Mental Health*

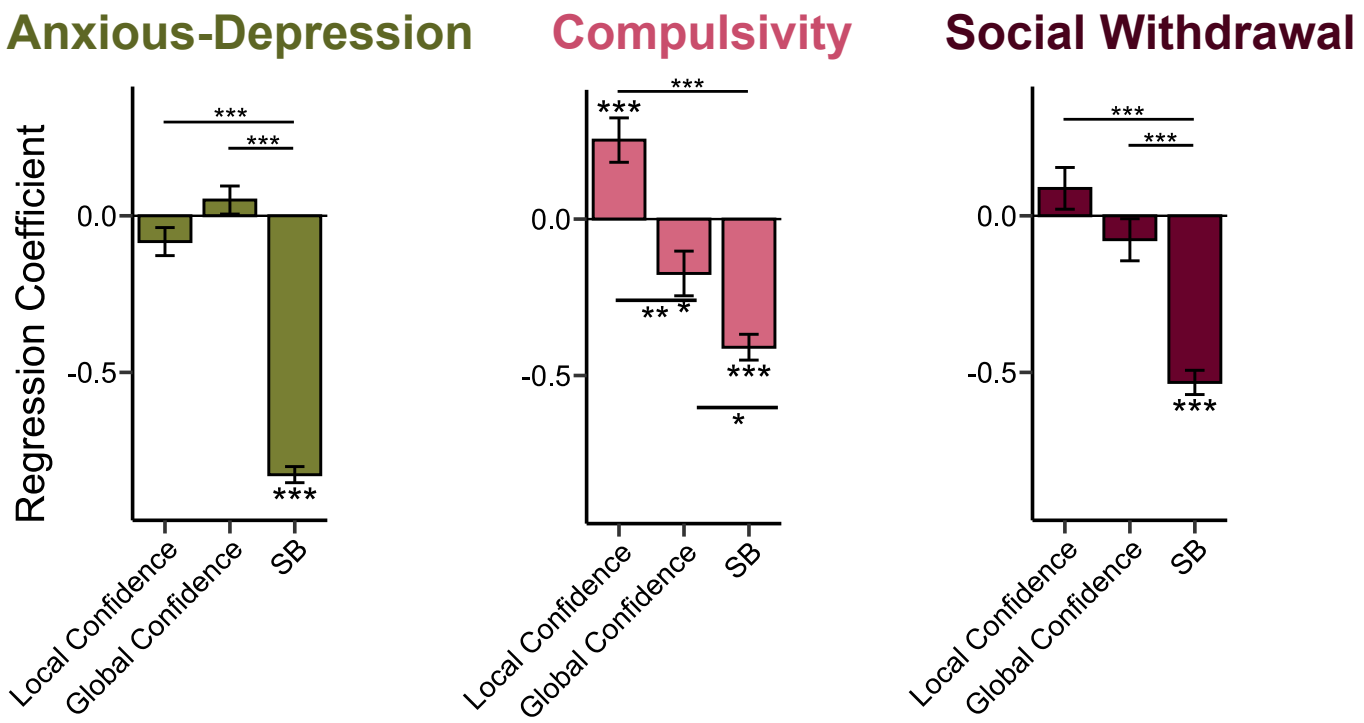
A hierarchy of metacognitive evaluations



Hoven, Denys, Luigjes, Rouault & van Holst (2023) *Nature Mental Health*

Global self-beliefs more strongly related to fluctuations in mental health symptoms

"Self-Belief" (SB) dimension: PC1 of autonomy, mastery, self-efficacy, self-esteem scale

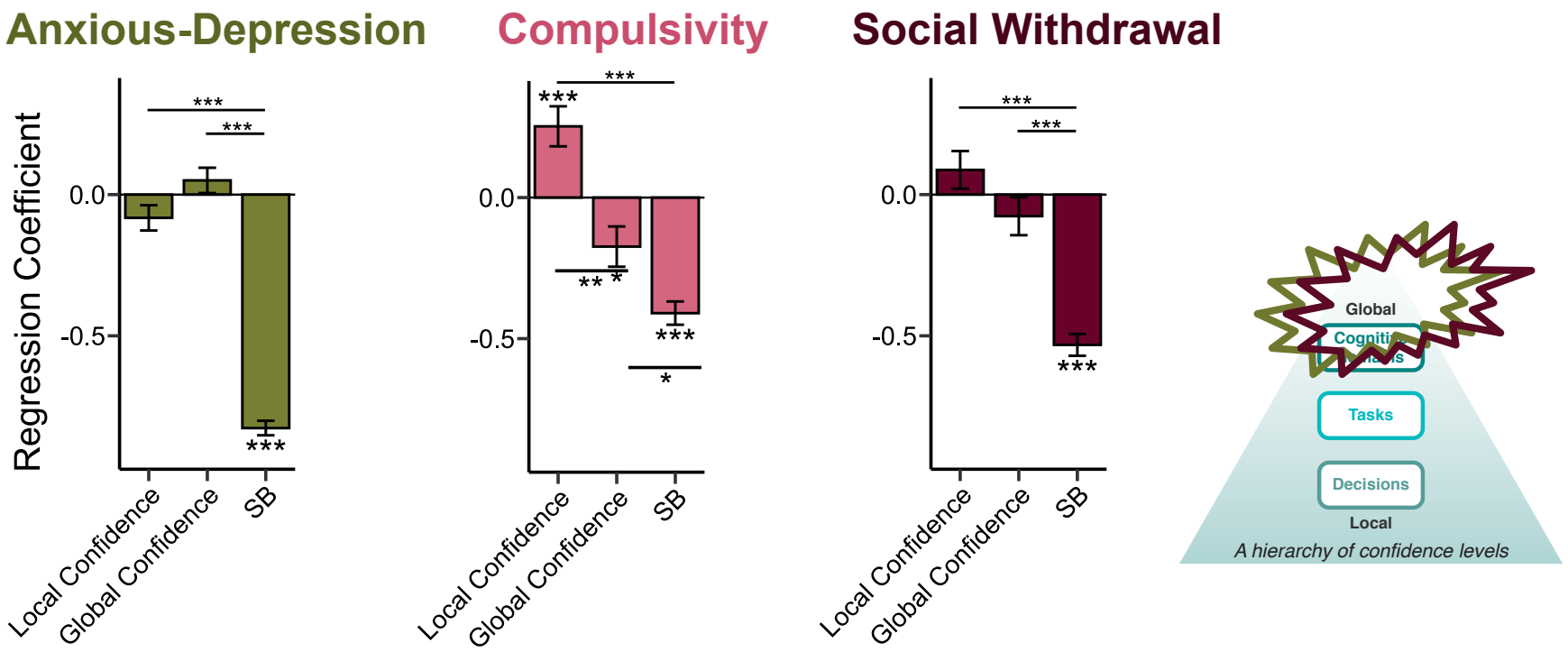


When in competition for explaining variance in symptom scores, global aspects of confidence contribute more ("SB" dimension)

Hoven, Denys, Luigjes, Rouault & van Holst (2023) *Nature Mental Health*

Global self-beliefs more strongly related to fluctuations in mental health symptoms

"Self-Belief" (SB) dimension: PC1 of autonomy, mastery, self-efficacy, self-esteem scale



When in competition for explaining variance in symptom scores, global aspects of confidence contribute more ("SB" dimension)

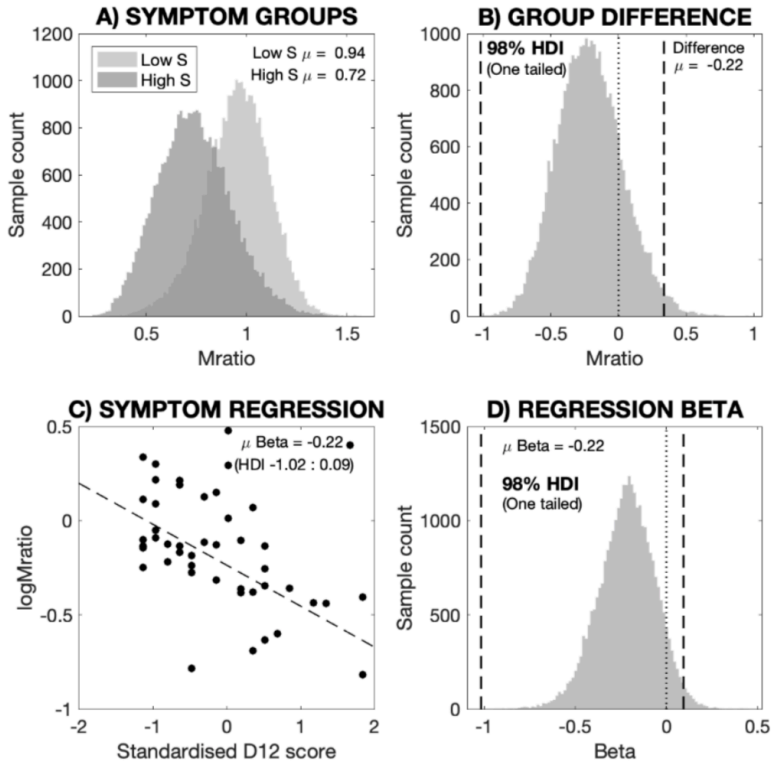
Seow*, Rouault* et al. (2022) *Biological Psychiatry*

« Psychiatry-informed » prior

Can we embed trait psychopathology directly into metacognitive efficiency estimation?

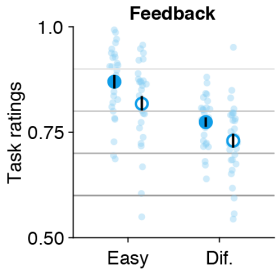
Hypothesis: trait characteristics will impact metacognitive behaviour: include as a co-regressor

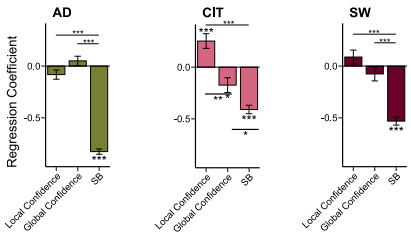
- Comparison of high and low symptom groups (asthma) in an interoception breathing task:

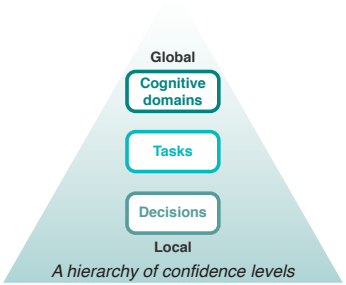


Moutoussis et al., (2019) *Frontiers*; Harrison et al., (2020) ***Biological Psychology***

Conclusions







- Self-beliefs formation may be related to global constructs such as self-esteem
- **Metacognitive profiles** relates to mental health symptom dimensions differently **at different hierarchical levels**
- **Global self-beliefs may be closer to functional and subjective symptoms** experienced by psychiatric and neurological patients

Acknowledgements

Dimensional psychiatry
C. GILLAN, T. SEOW
Trinity College Dublin

Self-esteem
G-J. WILL, **Univ. of Leiden**

Metacognition
S. FLEMING
P. DAYAN
Univ. College London

Transdiagnostic self-beliefs
M. HOVEN
R. van HOLST
J. LUIGJES
Univ. of Amsterdam

Thank you for your attention