

Higher-Order Awareness without First-Order Accuracy: Implications for Models of Metacognition

Ryan Scott^{1,3}, Zoltan Dienes^{1,3},
Adam Barrett^{2,3}, & Anil Seth^{2,3}

¹School of Psychology, University of Sussex

²School of Informatics, University of Sussex

³Sackler Centre for Consciousness Science

Metacognition - Knowing that we know

Typically operationalised using confidence

- Objective decision accuracy – demonstrates knowing
- Confidence-accuracy correlation – demonstrates metacognition

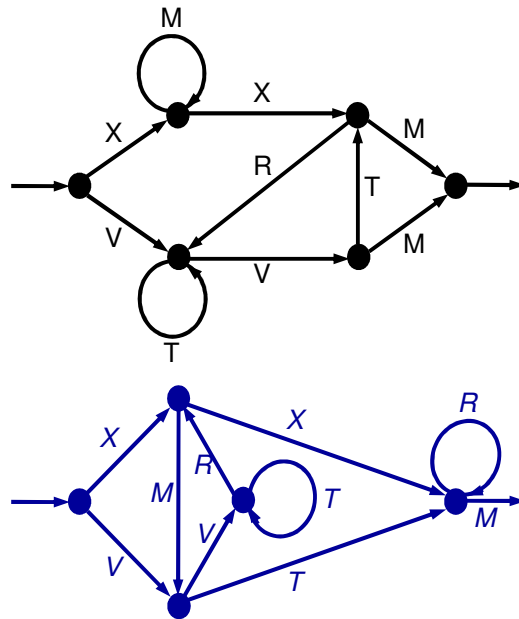
Signal detection has been applied to metacognition

- Lau (2008) proposed a SDT based framework for metacognition
- SDT measures (type I and II d-prime) widely used (Galvin,2003)

We sought an empirical test of the SDT based framework

- We analyse data from artificial grammar learning (AGL)
- Examine constraints inherent in the SDT framework
- Evaluate predictions relating to both type I and type II d'

Knowing and Metacognition in AGL



Training for Group A

XMMXM
VTTVTM
VVTRTVM
...

Testing for Group A and B

VTVTM
VTRRM
XXRVM
XXRRM
...

Training for Group B

XMTRM
VVRMTM
VTRRRRM
...

Judgements required

1. String grammaticality
(Decision accuracy)
2. Their confidence
(Metacognitive accuracy)

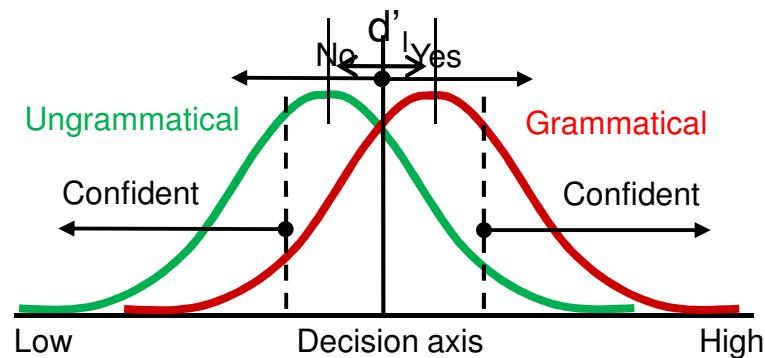
Typical results

1. Decision Accuracy variable
(60% - 70% is normal)
2. Metacognitive accuracy partial
(55% right without confidence)

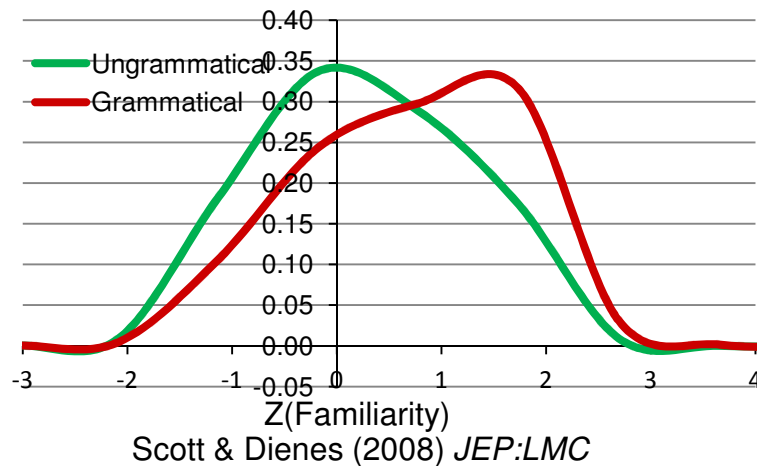
Signal Detection Theory applied to AGL

Type I d-prime (d'_{I})

Hit – Respond Grammatical when string IS Grammatical
 FA - Respond Grammatical when string IS NOT grammatical

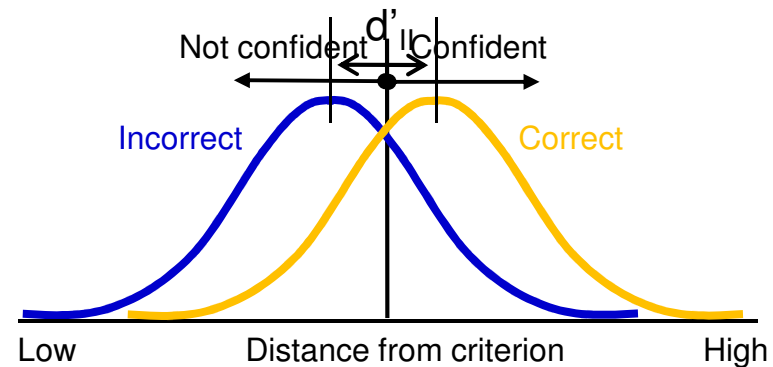


Equivalent distributions assuming responses based on subjective familiarity ratings (N = 384)

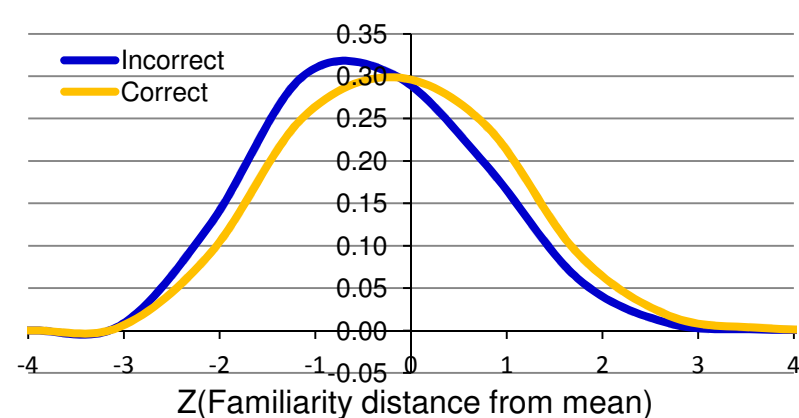


Type II d-prime (d'_{II})

Hit - Confident response when answer IS correct
 FA - Confident response when answer IS NOT correct

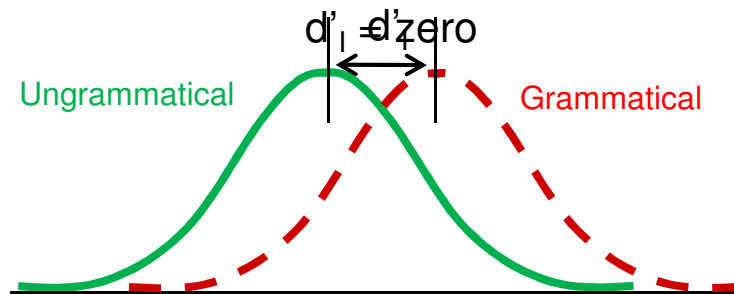


Equivalent distributions assuming symmetrical confidence bounds and mean criterion

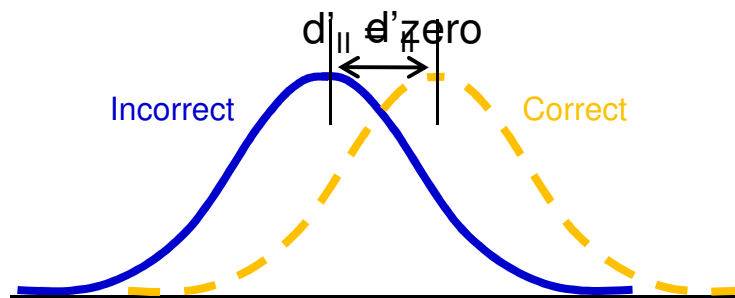


A constraint inherent to the SDT model

When type I d-prime is zero ...



... type II d-prime must also be zero



Evaluating metacognitive accuracy in the absence of decision accuracy

Data compiled from 8 standard AGL studies (N = 450)

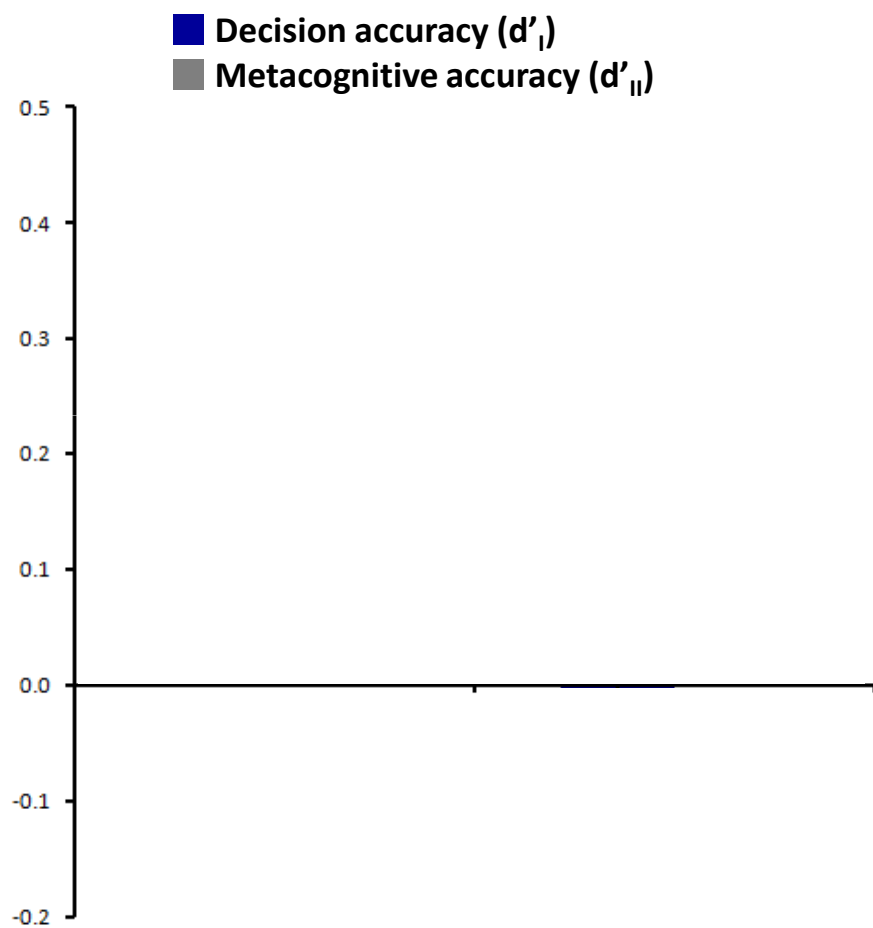
Participants provided decisions for 64 test strings

- Classified each string as grammatical or ungrammatical
- Rated their confidence in that judgment
- Reported decision strategy – Guess, Intuition, Rules, Memory

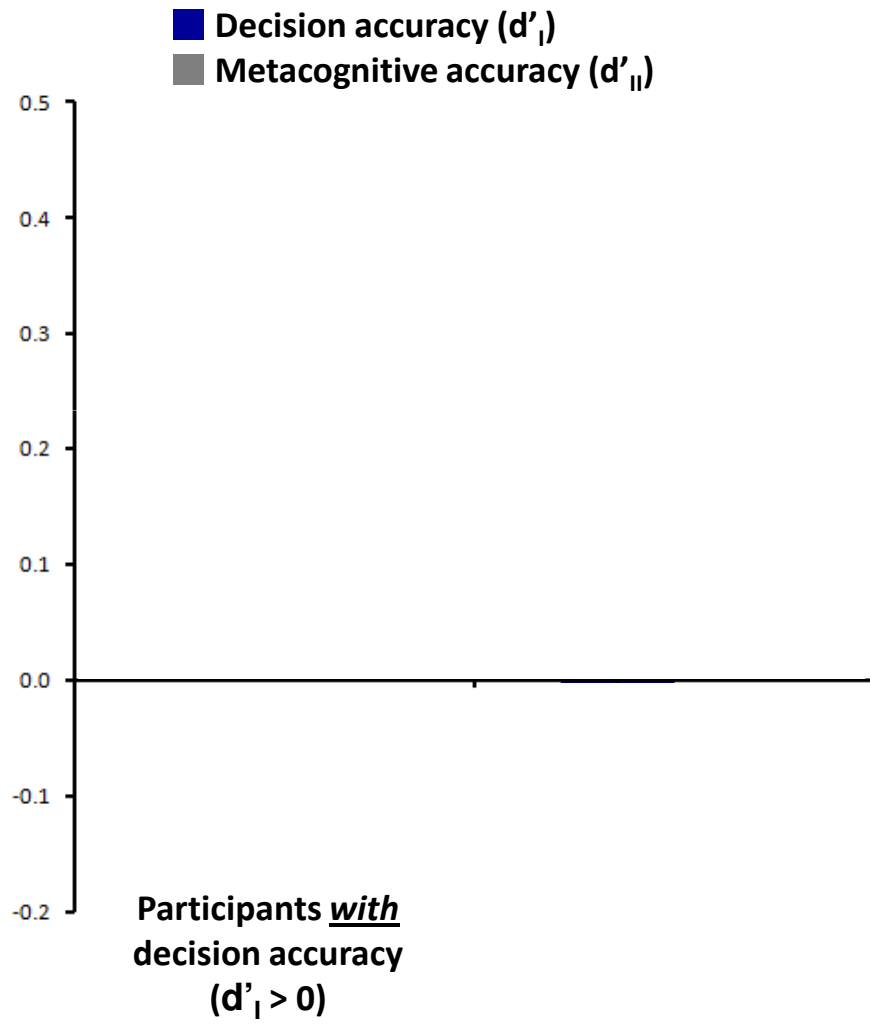
Selected according to decision accuracy while avoiding bias

- Categorised based on type I $d' \leq 0$ for the first $\frac{3}{4}$ of responses.
- Conducted analysis on the final $\frac{1}{4}$ of their responses.

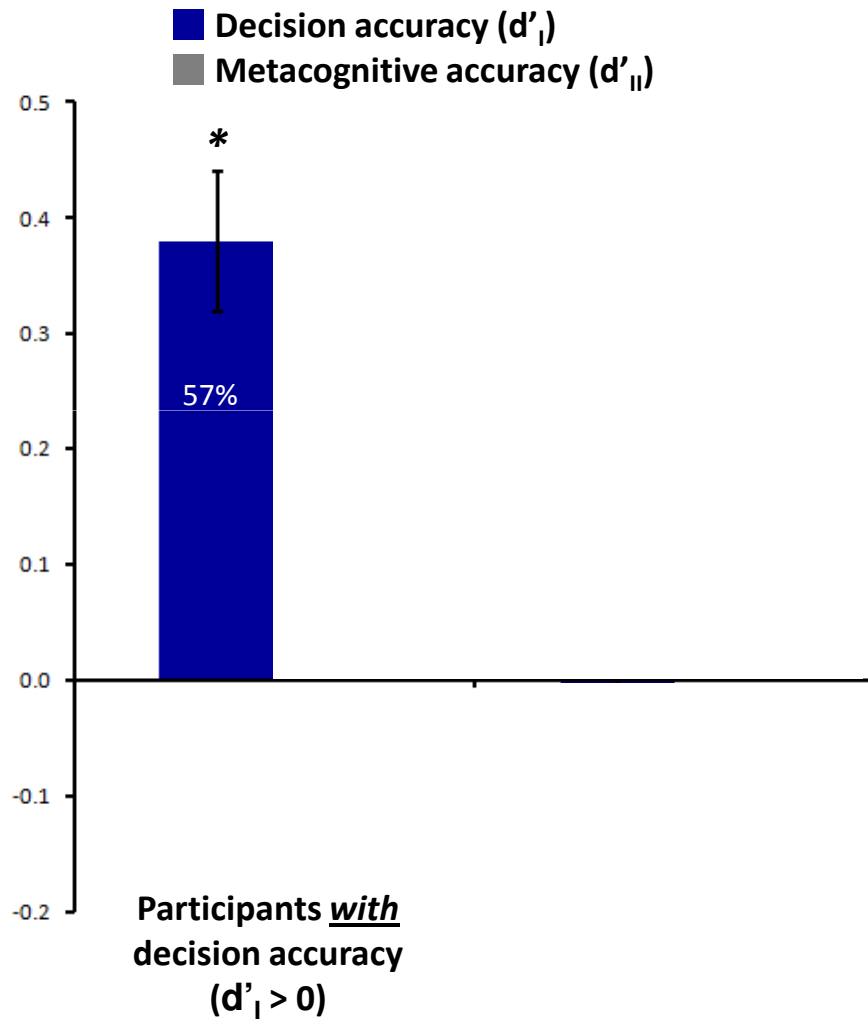
Results



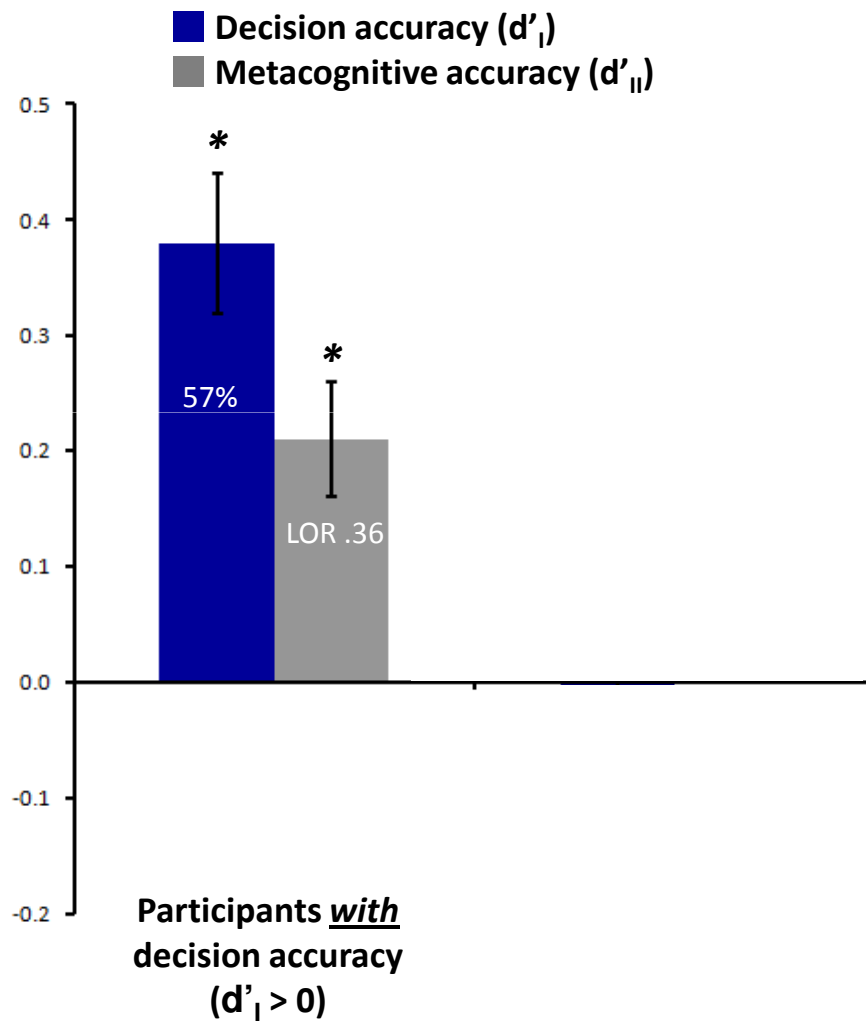
Results



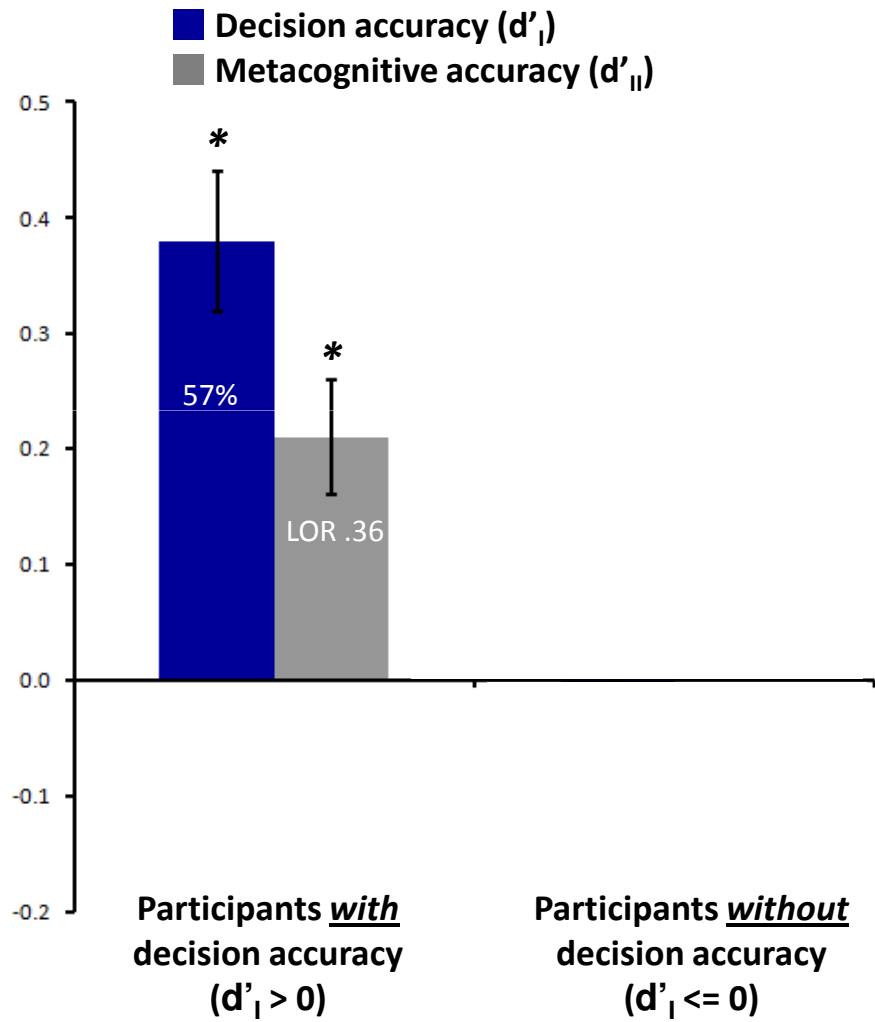
Results



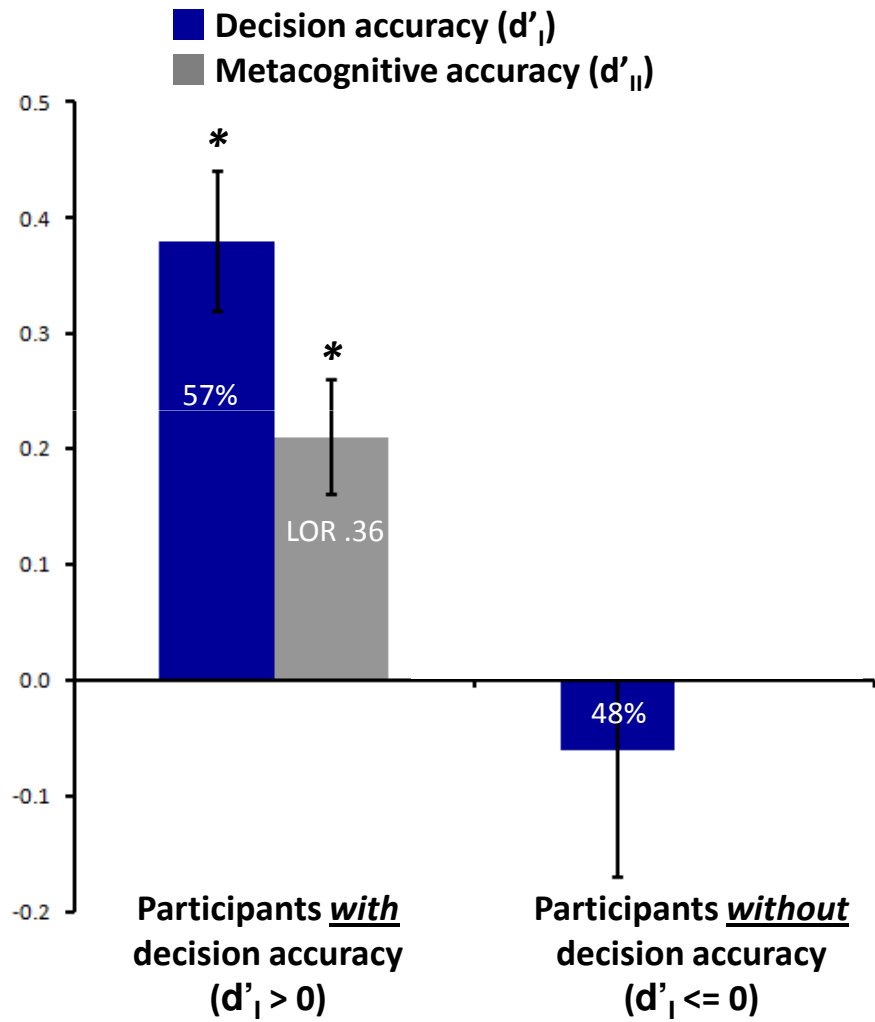
Results



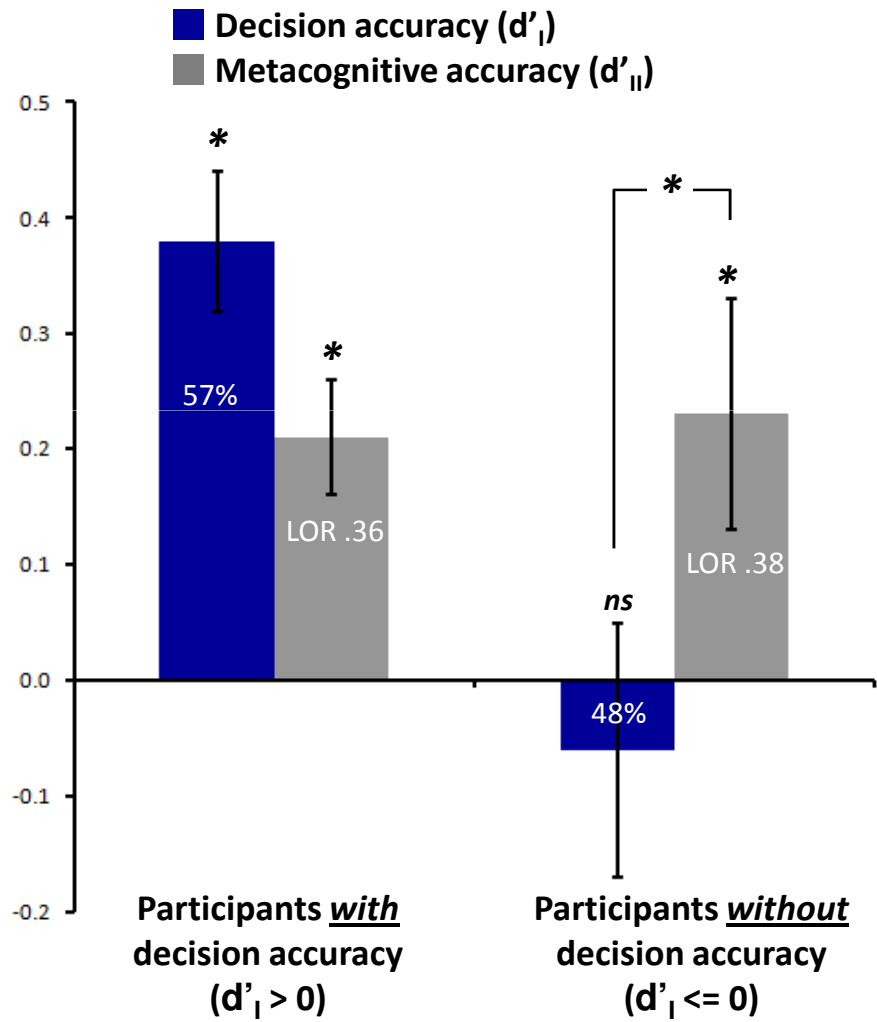
Results



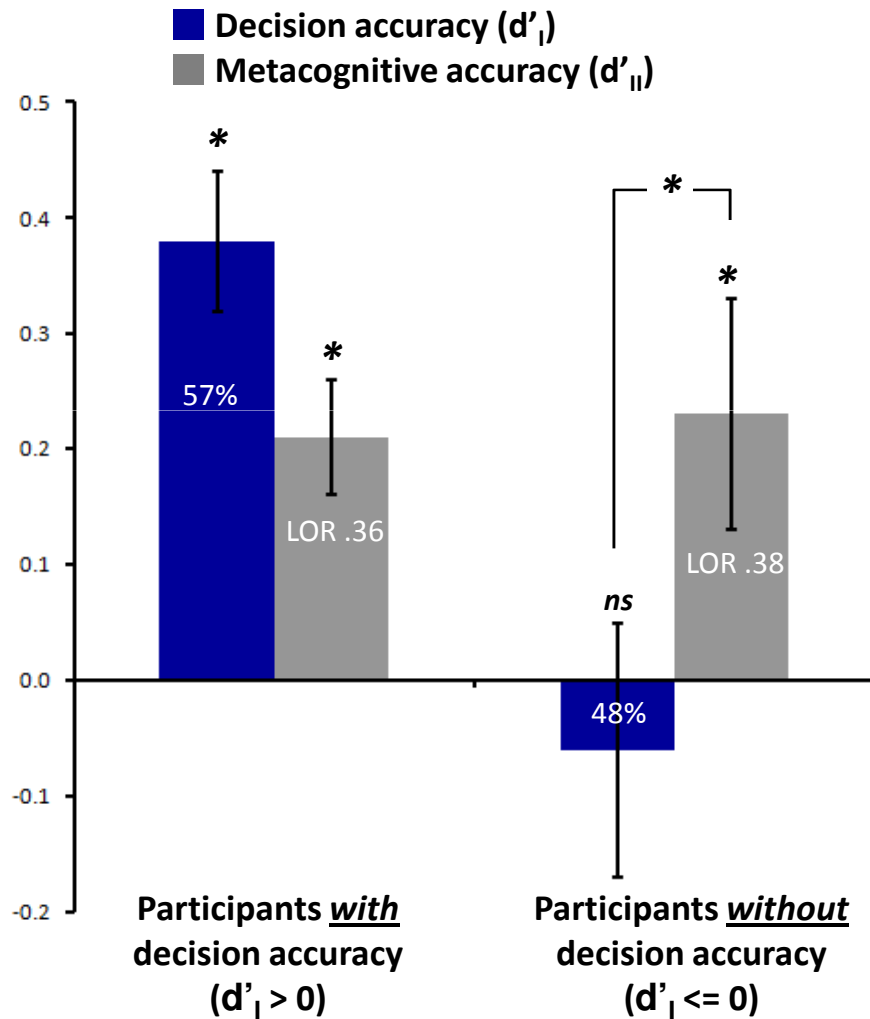
Results



Results



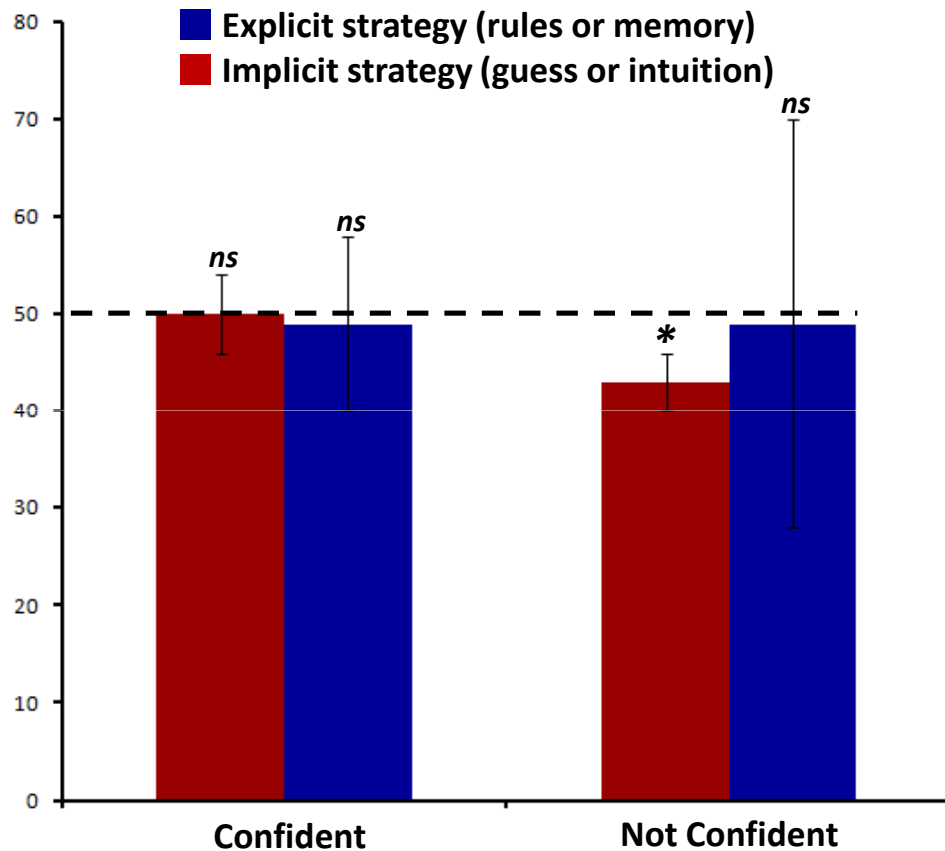
Results



Significant metacognitive accuracy in the absence of decision accuracy.

When type I d-prime is not significantly different from zero type II d-prime is significantly greater than zero.

What is driving the effect?



The relationship between confidence and accuracy is driven (largely) by having low confidence in wrong implicit decisions.

There would seem to be some implicit signal contributing to metacognitive accuracy

Thank you

Collaborators



Zoltan Dienes



Anil Seth



Adam Barrett

Funding and Support

