

# Regularised Neural Networks Mimic Human Insight

Anika T. Löwe<sup>1,2</sup>, Léo Touzo<sup>3</sup>, Paul S. Muhle-Karbe<sup>4,5,6</sup>, Andrew M. Saxe<sup>7,8,9</sup>, Christopher Summerfield<sup>\*\*</sup> & Nicolas W. Schuck<sup>1,2,10\*</sup>

<sup>1</sup>Max Planck Institute for Human Development, Berlin, Germany  
<sup>2</sup>Max Planck UCL Centre for Computational Psychiatry and Ageing Research, Berlin, Germany  
<sup>3</sup>Department of Physics, Ecole Normale Supérieure, Paris, France  
<sup>4</sup>Department of Experimental Psychology, University of Oxford, UK  
<sup>5</sup>School of Psychology, University of Birmingham, Birmingham, UK

<sup>6</sup>Centre for Human Brain Health, University of Birmingham, Birmingham, UK  
<sup>7</sup>Gatsby Computational Neuroscience Unit, University College London, London, UK  
<sup>8</sup>Sainsbury Wellcome Centre, University College London, London, UK  
<sup>9</sup>CI FAR Azrieli Global Scholar, CI FAR, Toronto, Canada  
<sup>10</sup>Institute of Psychology, Universität Hamburg, Hamburg, Germany

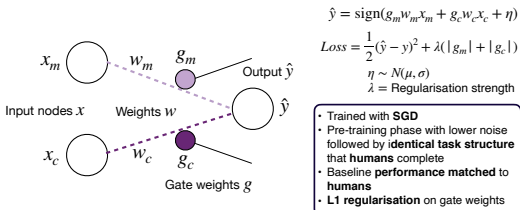
loewe@mpib-berlin.mpg.de  
 twitter.com/anika\_loewe

## BACKGROUND

- Insights, or aha-moments, are a remarkable phenomenon in human cognition
- Occur after a period of impasse [1], happen unusually abrupt [2] and only in some learners [3]
- Neural networks trained with SGD seem to imply that all learning is gradual
- We test insight-like learning dynamics (delay, suddenness, selectivity) in humans and L1-regularised gated neural networks

**Q: can insight-like learning arise naturally from gradual learning?**

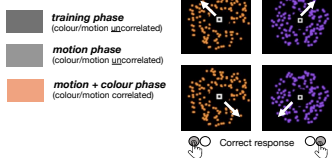
## NETWORK ARCHITECTURE



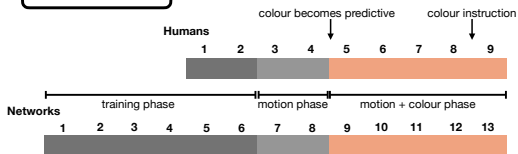
## TASK DESIGN

### Random Dot Motion Task

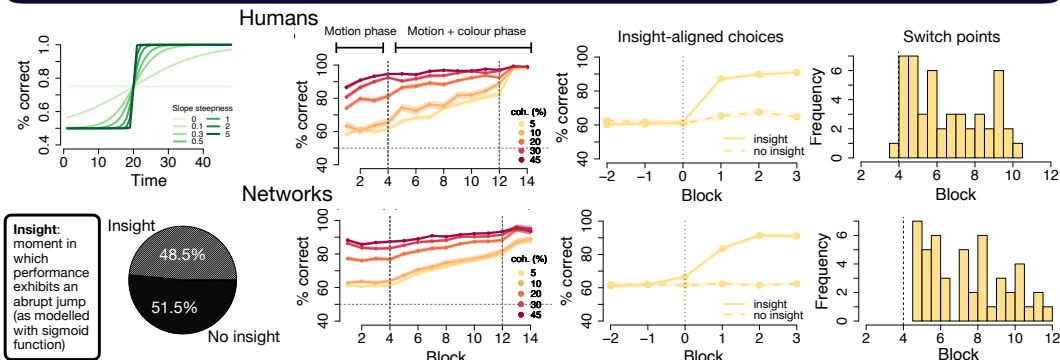
- 4 motion directions
- 2 colours
- 2 alternative forced choice
- Trial-wise binary feedback



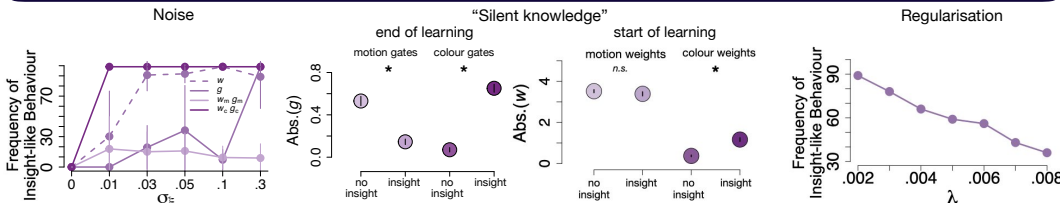
**Noise levels**  
 motion: varies in five steps  
 colour: no noise



## BEHAVIOUR



## LEARNING DYNAMICS



## CONCLUSION

- Humans tend to discover a hidden task regularity through insight, rather than gradually
- Neural networks with regularised gate modulation closely mimicked behavioural characteristics of human insights (delay, suddenness, selectivity)
- Insight-like behaviour in networks crucially depended on noise added to gradient updates, preceded by "silent knowledge" that is initially suppressed by regularised (attentional) gating
- We shed light on the computational origins of insights and suggest that they can arise naturally from gradual learning mechanisms

Preprint:

