Models of working memory aren’t working for me

For the Visual Memory Lab, University of Nottingham

21st March, 2025

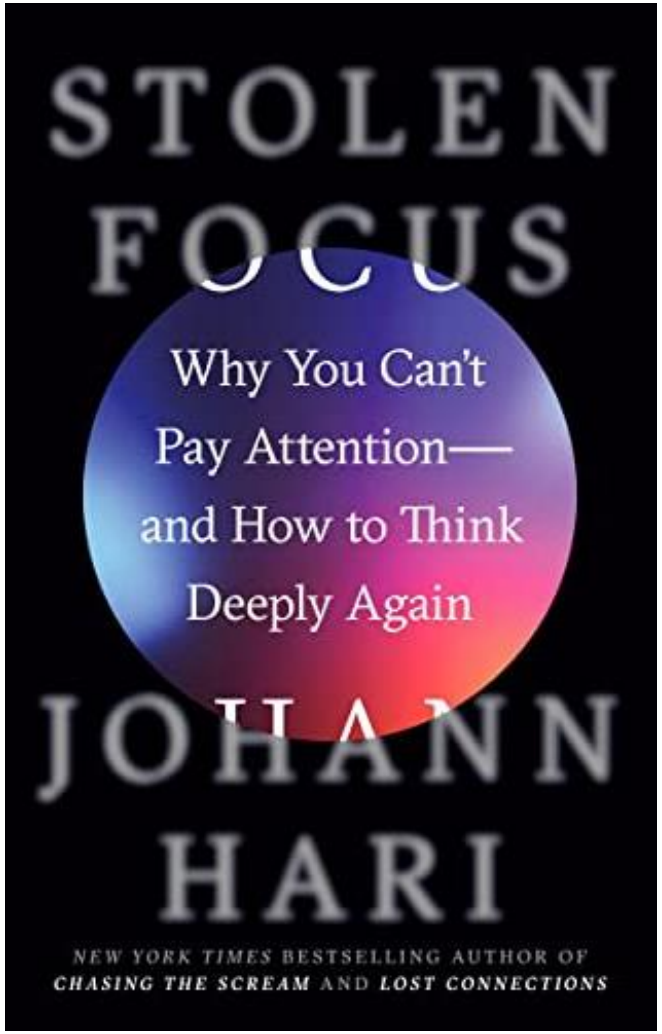
Adelaide University

Why study attention and memory?

Why study attention and memory?

- Our attention is very limited
- Therefore, our attention is *precious*
- We should want the right things to take up our attention!

Why study attention and memory?

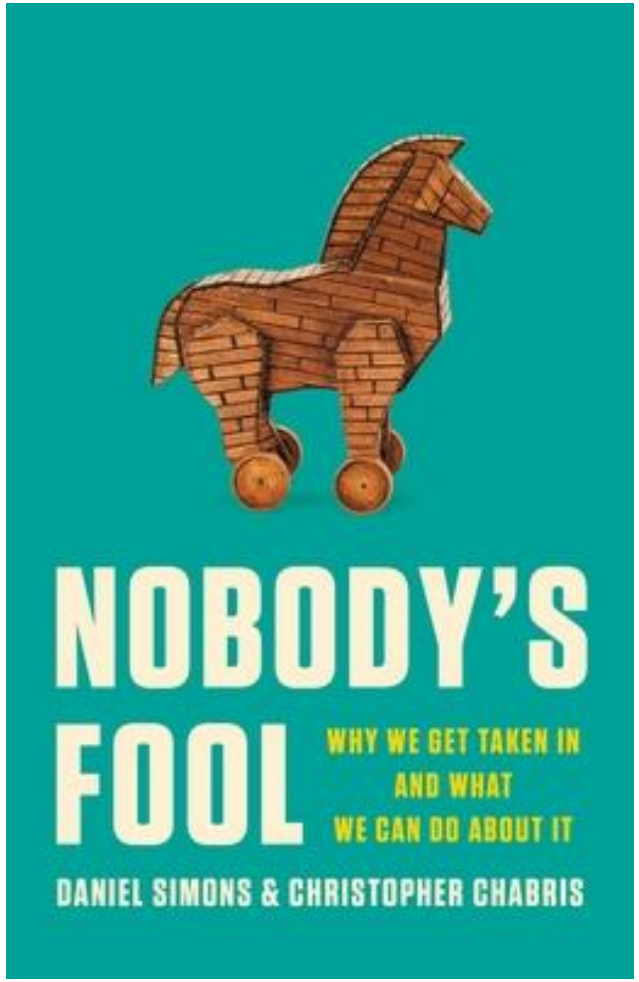


New York Times Bestseller, Book of the Year by Financial Times, etc.

Taps into the collective feeling that we are losing our ability to ***focus***

In my opinion, a very average book…

Why study attention and memory?



The researchers behind the "invisible gorilla" study!

We can be deceived when made to attend to the wrong things.

What is working memory?

What *is* visual working memory?

- “The system responsible for maintaining visual information in a state of heightened accessibility for ongoing perception and cognition.”

- This same definition could also describe visual **attention**
  - Perhaps also visual imagery, psychological introspection

What *is* visual working memory?

- Many subtly different definitions:

The many faces of working memory and short-term storage

Nelson Cowan

Psychonomic Bulletin & Review **24**, 1158–1170 (2017) | Cite this article

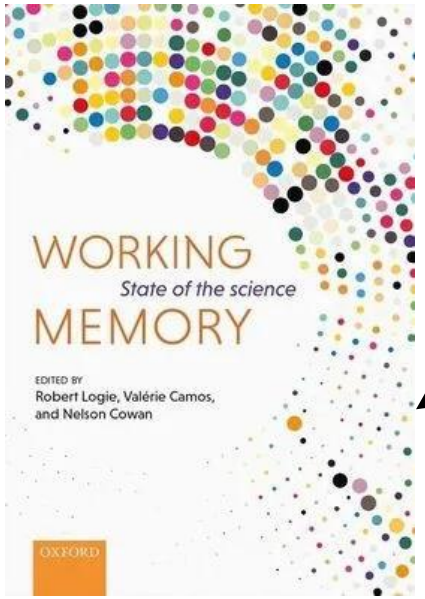
It has become clearer to me that a major source of confusion is that researchers use different definitions of the malleable and useful concept of WM. We do not seem to be converging on a common definition of the term. Others also have

What *is* visual working memory?

14

Integrating Theories of Working Memory

Robert H. Logie, Clément Belletier, and Jason M. Doherty



First published in late 2020

- "We argue that many of these differences reflect different research questions, different levels of explanation, differences in how participants perform their assigned tasks in different laboratories, rather than fundamental theoretical adversity"

Robert H. Logie, Clément Belletier, and Jason M. Doherty, Integrating Theories of Working Memory In: Working Memory. Edited by: Robert H. Logie, Valérie Camos, and Nelson Cowan, Oxford University Press (2021). © Oxford University Press. DOI: 10.1093/oso/9780198842286.003.0014

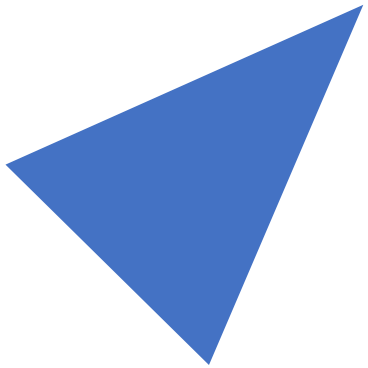
What *is* visual working memory?

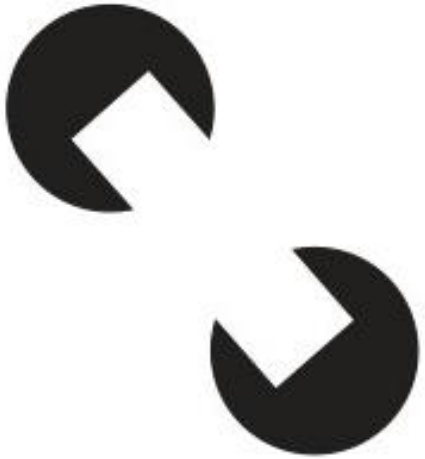
- “The system responsible for maintaining visual information in a state of heightened accessibility for ongoing perception and cognition.”

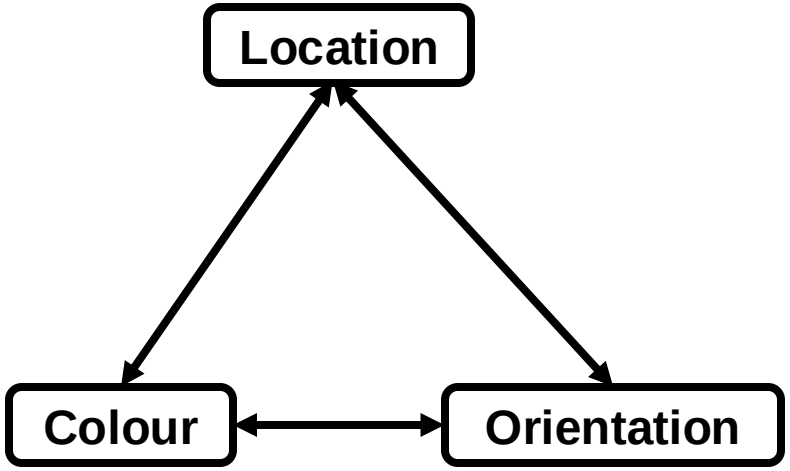
- This same definition could also describe visual **attention**
  - Perhaps also visual imagery, psychological introspection

- The core question: **How is information represented in mind?**

Representations in the mind







Constituents?
Illusory objects?

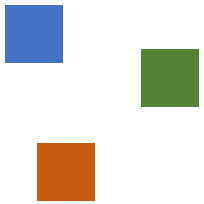
Memories across space and time?

What *is* visual working memory?

Object-based theory

"slot models"

(Luck and Vogel, 1997;
Zhang and Luck, 2008)



Feature-based theory

"*resource models*"

(Alvarez and Cavanagh, 2004;
Wilken and Ma, 2004)

Luck, S. J., & Vogel, E. K. (1997). https://doi.org/10.1038/36846
Zhang, W., & Luck, S. J. (2008). https://doi.org/10.1038/nature06860
Alvarez, G. A., & Cavanagh, P. (2004). https://doi.org/10.1111/j.0963-7214.2004.01502006.x
Wilken, P., & Ma, W. J. (2004). https://doi.org/10.1167/4.12.11

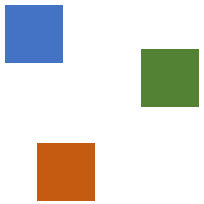
What *is* visual working memory?

- An enduring theoretical framework has been

Object-based theory

"slot models"

(Luck and Vogel, 1997;
Zhang and Luck, 2008)



versus

Feature-based theory

"*resource models*"

(Alvarez and Cavanagh, 2004;
Wilken and Ma, 2004)

Luck, S. J., & Vogel, E. K. (1997). https://doi.org/10.1038/36846
Zhang, W., & Luck, S. J. (2008). https://doi.org/10.1038/nature06860
Alvarez, G. A., & Cavanagh, P. (2004). https://doi.org/10.1111/j.0963-7214.2004.01502006.x
Wilken, P., & Ma, W. J. (2004). https://doi.org/10.1167/4.12.11

A *theory crisis* in psychological science

- An understated precursor to the *reproducibility crisis* may be the lack of coordinated theoretical development
  - An over-reliance on the hypothetico-deductive method (e.g. null hypothesis significance testing) for inferences
    - Questionable research practices (QRPs): *p*-hacking, HARKing, data manipulation, etc.

Borsboom D. (2013, November 20). Theoretical amnesia. *Center for Open Science*

Borsboom, D., van der Maas, H. L., Dalege, J., Kievit, R. A., & Haig, B. D. (2021). Theory construction methodology: A practical framework for building theories in psychology. *Perspectives on Psychological Science*, *16*(4), 756-766.

Oberauer K., Lewandowsky S. (2019). Addressing the theory crisis in psychology. *Psychonomic Bulletin & Review*, 26, 1596–1618.

Maatman, F. O. (2021). Psychology's theory crisis, and why formal modelling cannot solve it. *PsyArXiv*

Meehl P. E. (1978). Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald, and the slow progress of soft psychology. *Journal of Consulting and Clinical Psychology*, 46, 806–834.

Playing *20 questions* with nature

- It is often assumed that…

Newell, A. (1973). You can't play 20 questions with nature and win: Projective comments on the papers of this symposium.

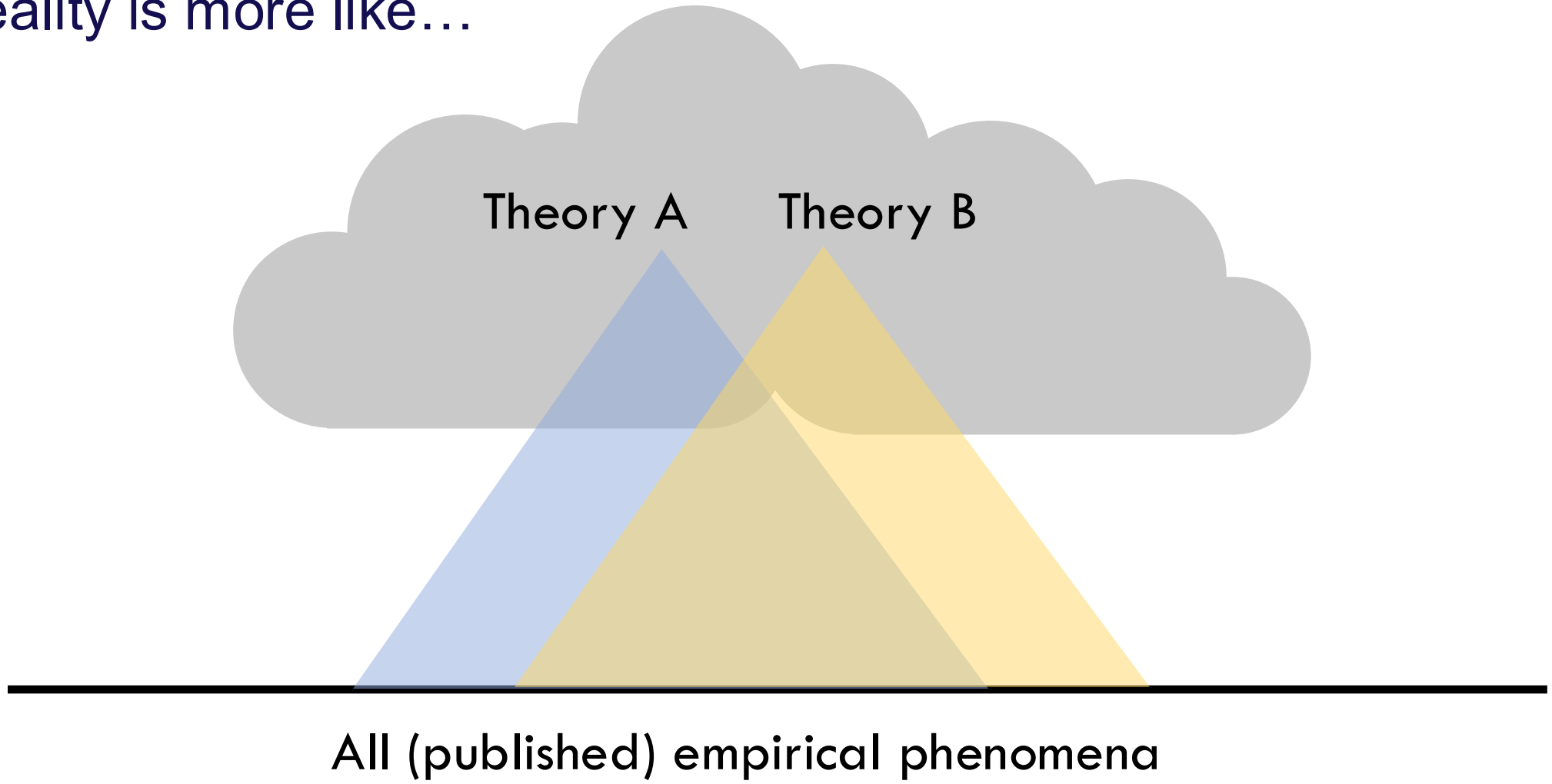
Playing *20 questions* with nature

- It is often assumed that…

Newell, A. (1973). You can't play 20 questions with nature and win: Projective comments on the papers of this symposium.

Playing *20 questions* with nature

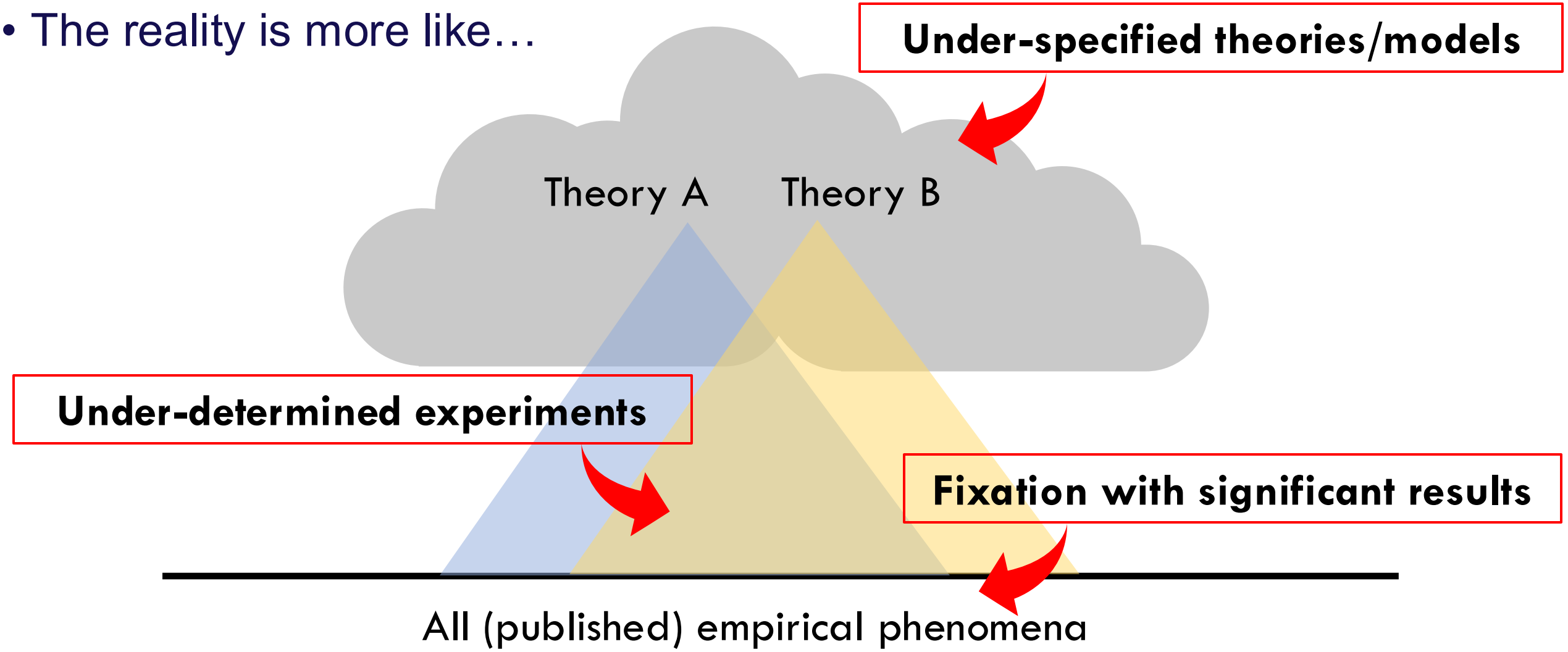
- The reality is more like…



Newell, A. (1973). You can't play 20 questions with nature and win: Projective comments on the papers of this symposium.

Playing *20 questions* with nature

- The reality is more like…



Newell, A. (1973). You can't play 20 questions with nature and win: Projective comments on the papers of this symposium.

A *theory crisis* in psychological science

- An understated precursor to the *reproducibility crisis* may be the lack of coordinated theoretical development
  - An over-reliance on the hypothetico-deductive method (e.g. null hypothesis significance testing) for inferences
    - Questionable research practices (QRPs): *p*-hacking, HARKing, data manipulation, etc.
  - Under-specified theories with under-determined experimental designs
    - *Ad hoc* changes in models, straw-man of competing models, blunt instruments of measurement
  - Overgeneralization of a theory or model to all related phenomena or empirical conditions
    - A lack of intellectual humility…

Borsboom D. (2013, November 20). Theoretical amnesia. *Center for Open Science*

Borsboom, D., van der Maas, H. L., Dalege, J., Kievit, R. A., & Haig, B. D. (2021). Theory construction methodology: A practical framework for building theories in psychology. *Perspectives on Psychological Science*, *16*(4), 756-766.

Oberauer K., Lewandowsky S. (2019). Addressing the theory crisis in psychology. *Psychonomic Bulletin & Review*, 26, 1596–1618.

Maatman, F. O. (2021). Psychology's theory crisis, and why formal modelling cannot solve it. *PsyArXiv*

Meehl P. E. (1978). Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald, and the slow progress of soft psychology. *Journal of Consulting and Clinical Psychology*, 46, 806–834.

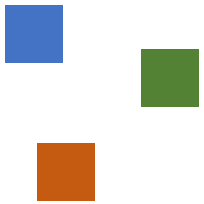
What *is* visual working memory?

- An enduring theoretical framework has been

Object-based theory

"slot models"

(Luck and Vogel, 1997;
Zhang and Luck, 2008)



versus

Feature-based theory

"*resource models*"

(Alvarez and Cavanagh, 2004;
Wilken and Ma, 2004)

Luck, S. J., & Vogel, E. K. (1997). https://doi.org/10.1038/36846
Zhang, W., & Luck, S. J. (2008). https://doi.org/10.1038/nature06860
Alvarez, G. A., & Cavanagh, P. (2004). https://doi.org/10.1111/j.0963-7214.2004.01502006.x
Wilken, P., & Ma, W. J. (2004). https://doi.org/10.1167/4.12.11

How do we address the theory crisis?

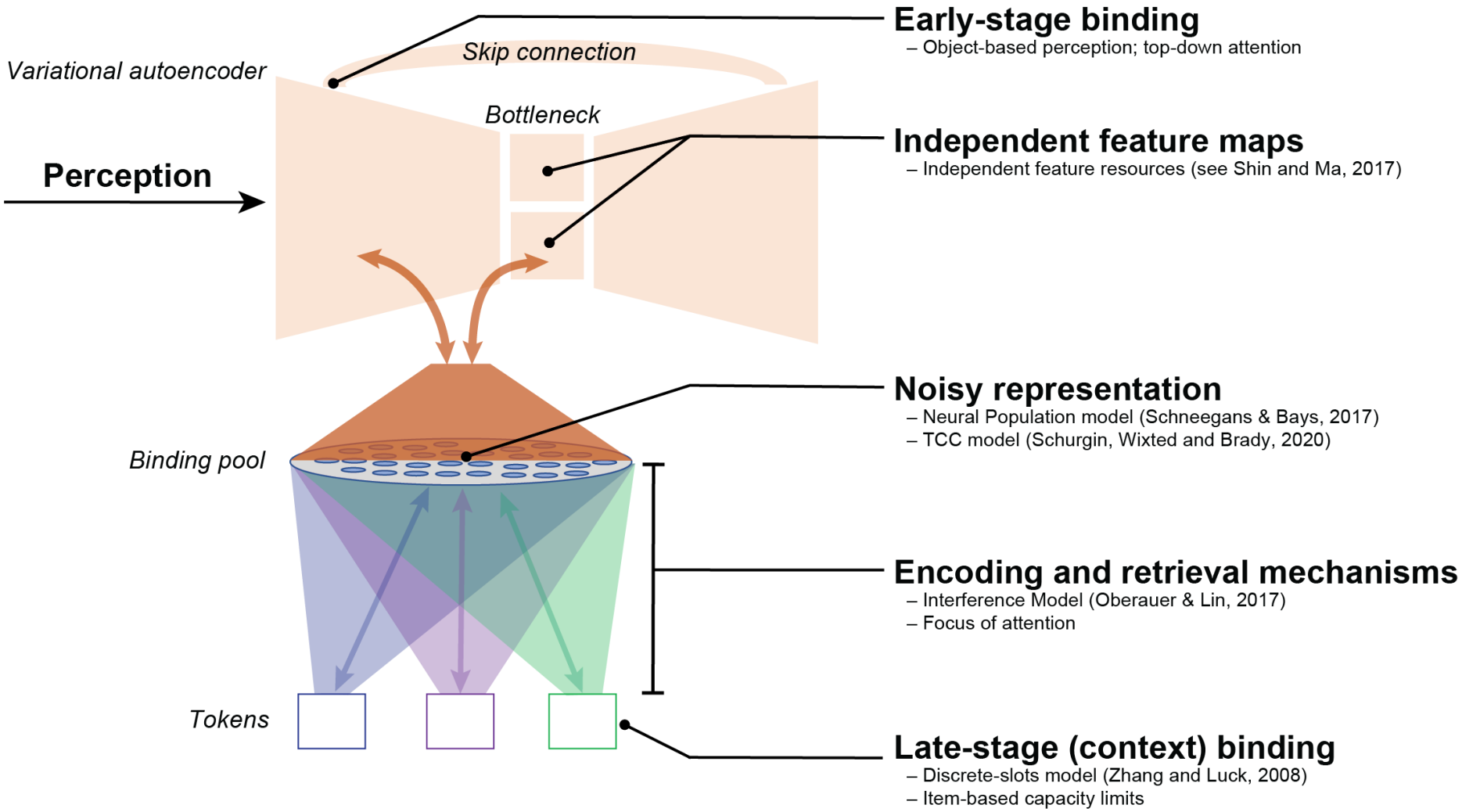
Can we bring these models into accordance?

Towards a model-centric science

- We need to move away from dualistic experiments and a results-oriented science towards a model-centric science
- We need more theory development
  - Repeating and detailing the phenomena that we hope to explain
  - Integrating various empirical results and models
  - Clear specification of theories and models and how they relate to the phenomena
  - Careful generalization of current models (i.e. practicing intellectual humility)
  - Better thought-out methods and measures
  - Rigorous design of experiments to truly test hypotheses

Devezer, B., & Buzbas, E. O. (2023, April 17). Rigorous exploration in a model-centric science via epistemic iteration. https://doi.org/10.31222/osf.io/qe46u

Presenting a theory map for visual working memory

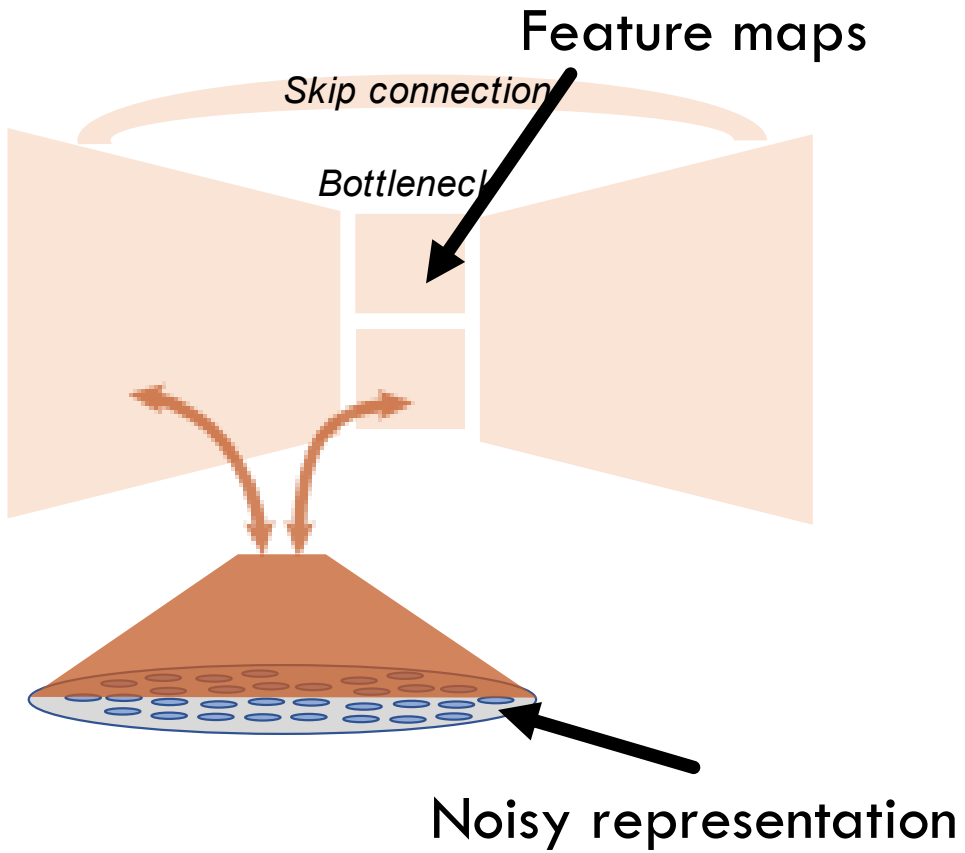


Hedayati, S., O'Donnell, R. E., & Wyble, B. (2022). A model of working memory for latent representations. *Nature Human Behaviour*, *6*(5), 709-719.

Binding pool as a locus for feature-based ideas

- Independent feature layers project into the binding pool (Shin and Ma, 2017)
  - But early-stage object-based attention may also be in play

- Noisy representations in VWM are well-captured by neural population and signal detection accounts (Bays, 2014; Schurgin et al., 2020)

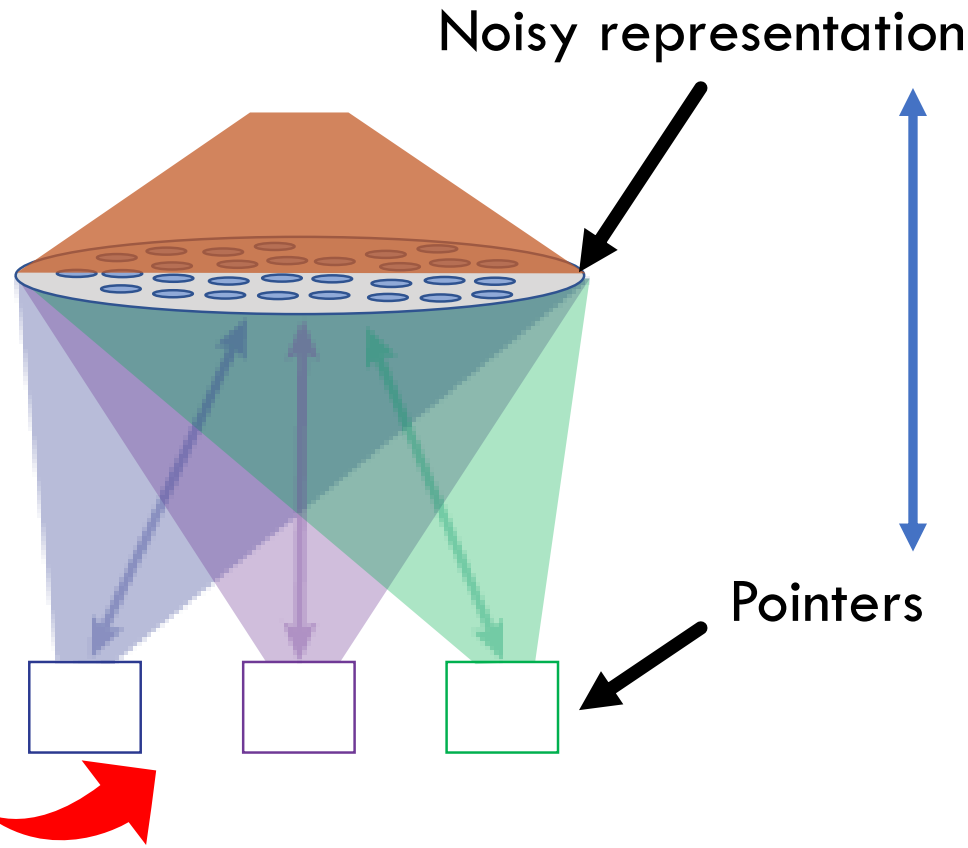


Bays, P. M. (2014). Noise in neural populations accounts for errors in working memory. *Journal of Neuroscience*, *34*(10), 3632-3645.
Schurgin, M. W., Wixted, J. T., & Brady, T. F. (2020). Psychophysical scaling reveals a unified theory of visual memory strength. *Nature human behaviour*, *4*(11), 1156-1172.
Shin, H., & Ma, W. J. (2017). Visual short-term memory for oriented, colored objects. Journal of Vision, 17(9), 12-12.

Tokens as a locus for object-based ideas

- Content-independent pointers
  - Like *FINSTs* or *Object Files* (Pylyshyn, 1989; Kahneman et al., 1992)

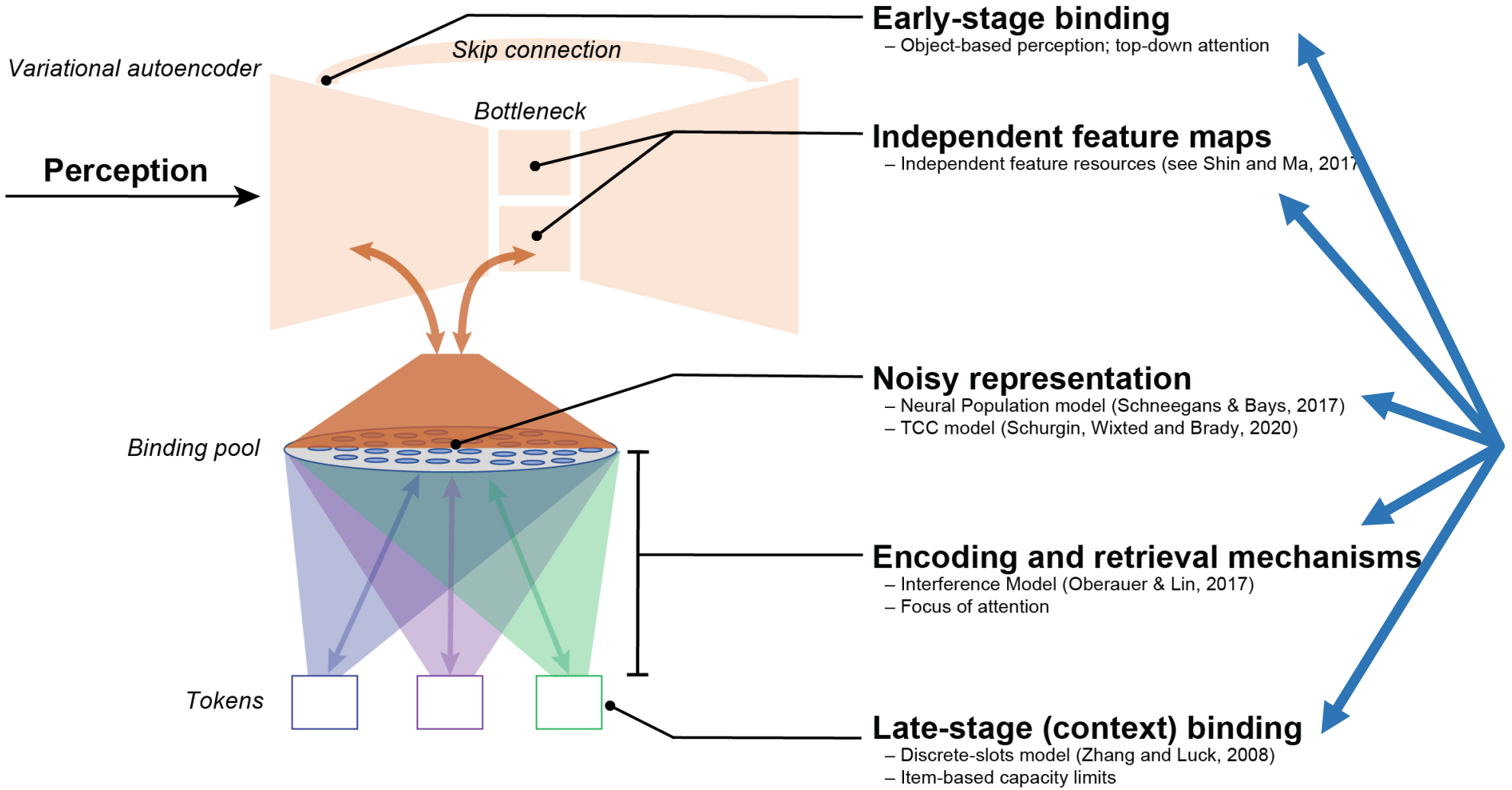
- Evidence for a neural signature that indexes VWM load and generalizes across feature content (Thyer et al., 2022; Balaban et al., 2019)



New conception of working memory as a very late-stage of encoding and selection

Pylyshyn, Z. (1989). The role of location indexes in spatial perception: A sketch of the FINST spatial-index model. *Cognition*, *32*(1), 65-97.
Kahneman, D., Treisman, A., & Gibbs, B. J. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive psychology*, *24*(2), 175-219.
Thyer, W., Adam, K. C., Diaz, G. K., Velazquez Sanchez, I. N., Vogel, E. K., & Awh, E. (2022). Storage in visual working memory recruits a content-independent pointer system. *Psychological Science*, *33*(10), 1680-1694.
Balaban, H., Drew, T., & Luria, R. (2019). Neural evidence for an object-based pointer system underlying working memory. *cortex*, *119*, 362-372.

Presenting a theory map for visual working memory



Empirical phenomena like capacity limits or biases in representations or anything can occur and interact at any of these levels

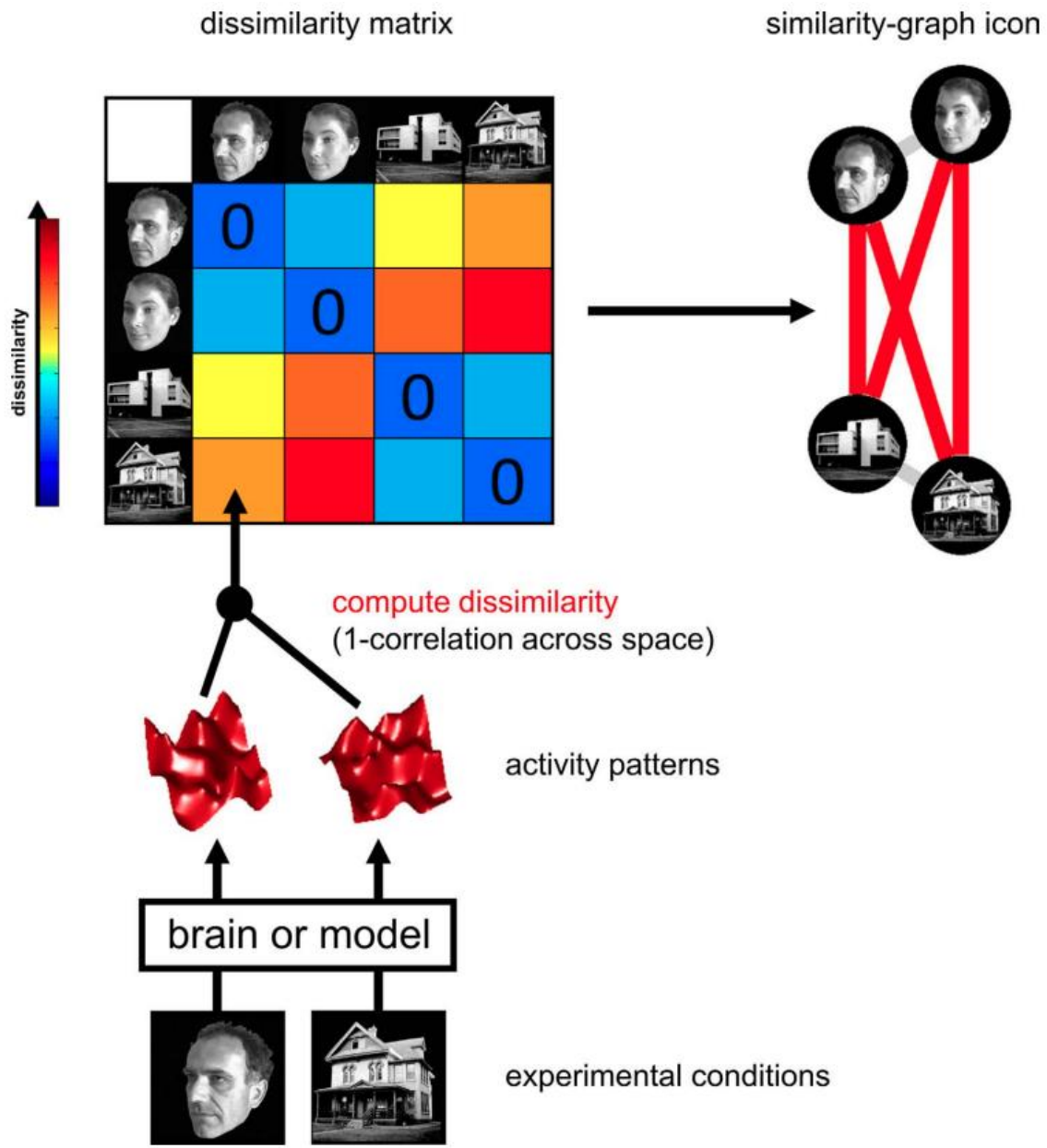
Hedayati, S., O'Donnell, R. E., & Wyble, B. (2022). A model of working memory for latent representations. *Nature Human Behaviour*, *6*(5), 709-719.

How does a theory map help?

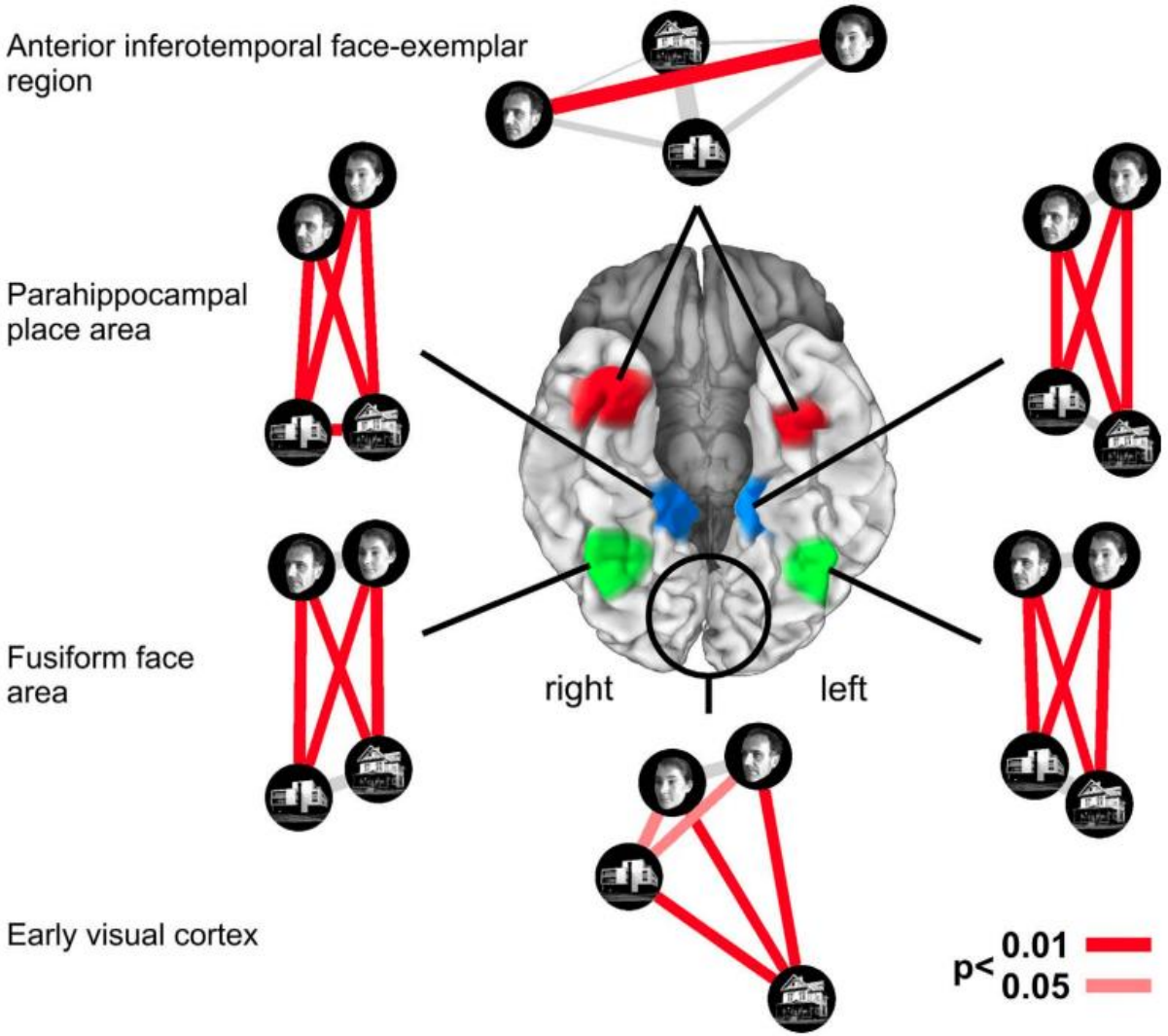
- Provides a common core language and framework to discuss theories, models, and phenomena
  - Reveals hidden intuitions
  - Prevents misunderstandings from varying definitions
  - Better specifies connection between models and phenomena
  - Reduces straw-man of various positions
  - Discourages a dualistic framework for experimental design
  - Initiates better determined model comparisons and definitive empirical tests

- Inspires theory development
  - Promotes counterinduction (the use and development of others' models)
  - Encourages slow science from better thought-out studies

A brief overview of multivariate decoding

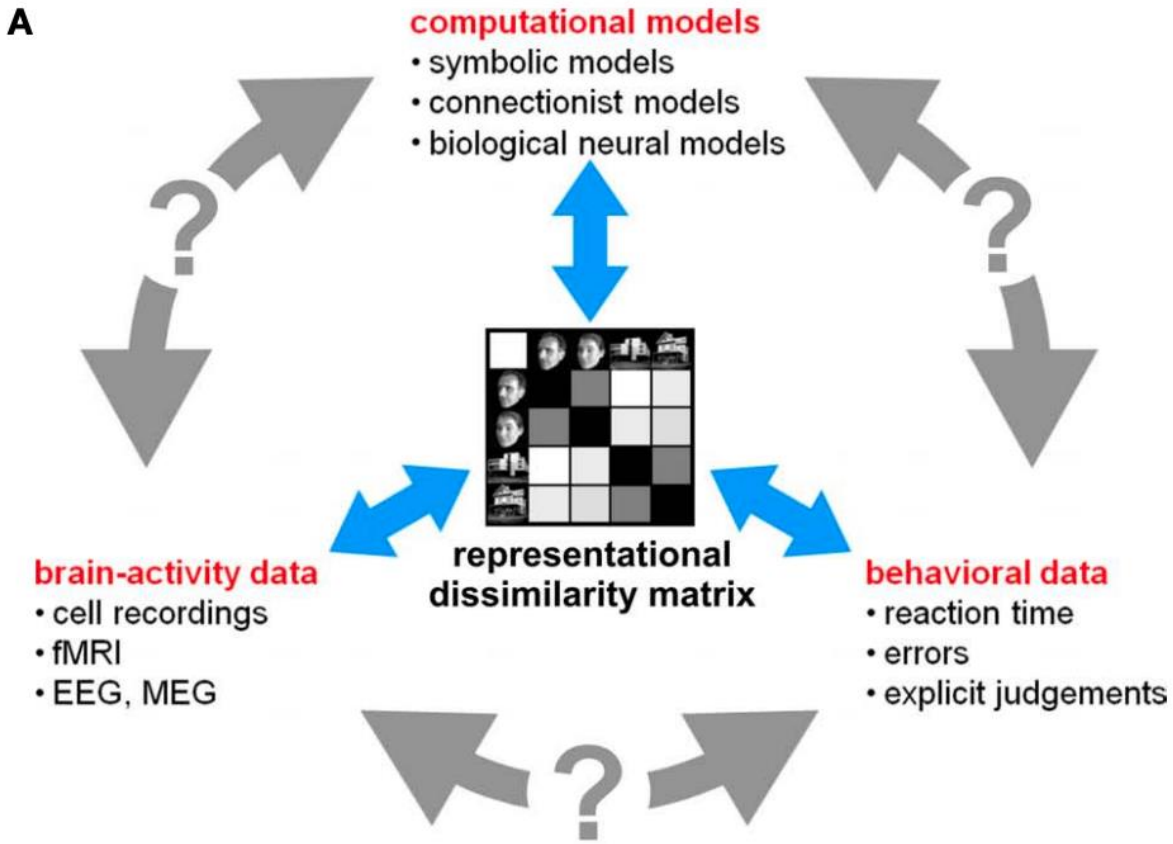


Kriegeskorte, N., Mur, M., & Bandettini, P. A. (2008). Representational similarity analysis-connecting the branches of systems neuroscience. *Frontiers in systems neuroscience*, *2*, 249.

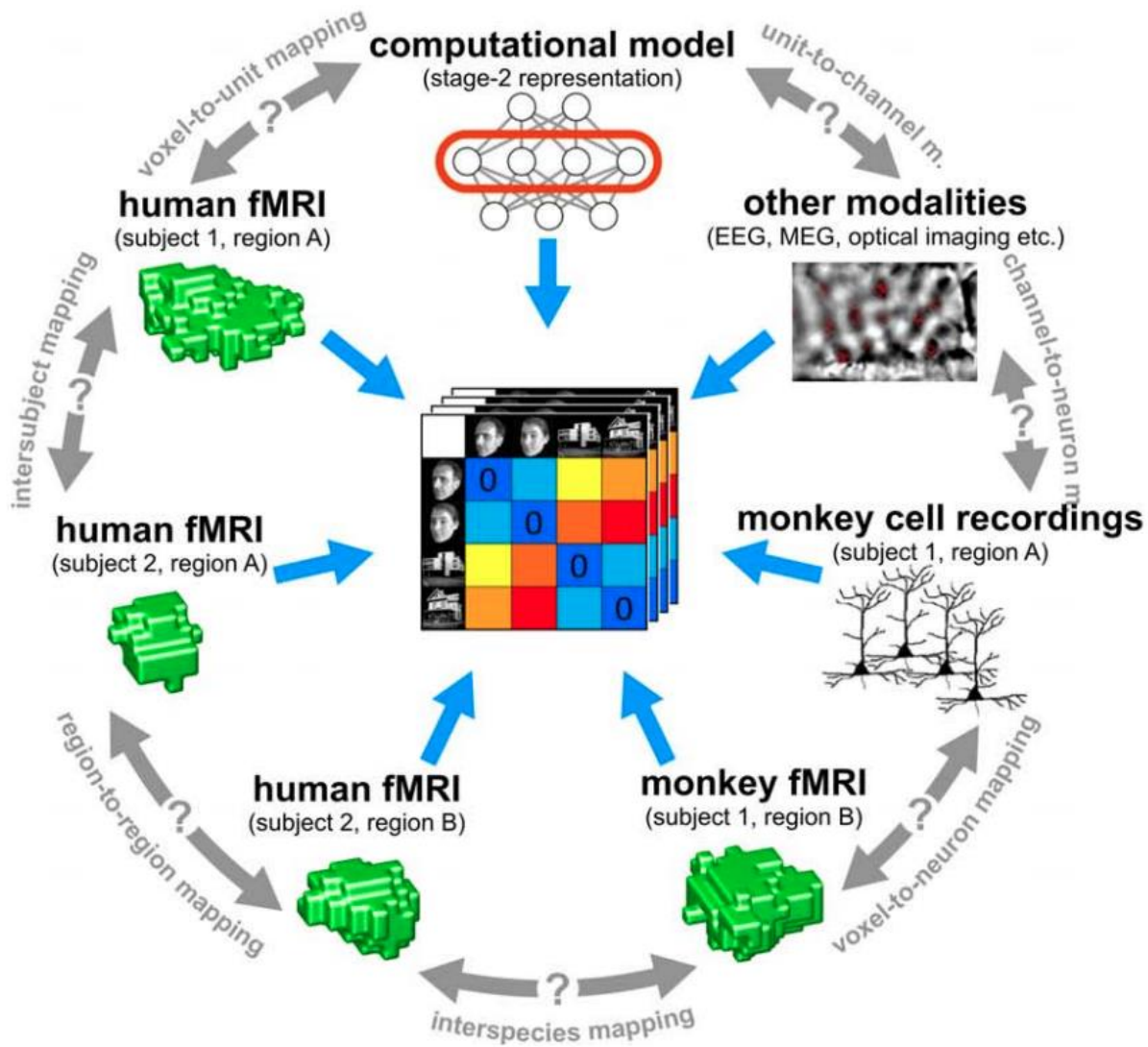


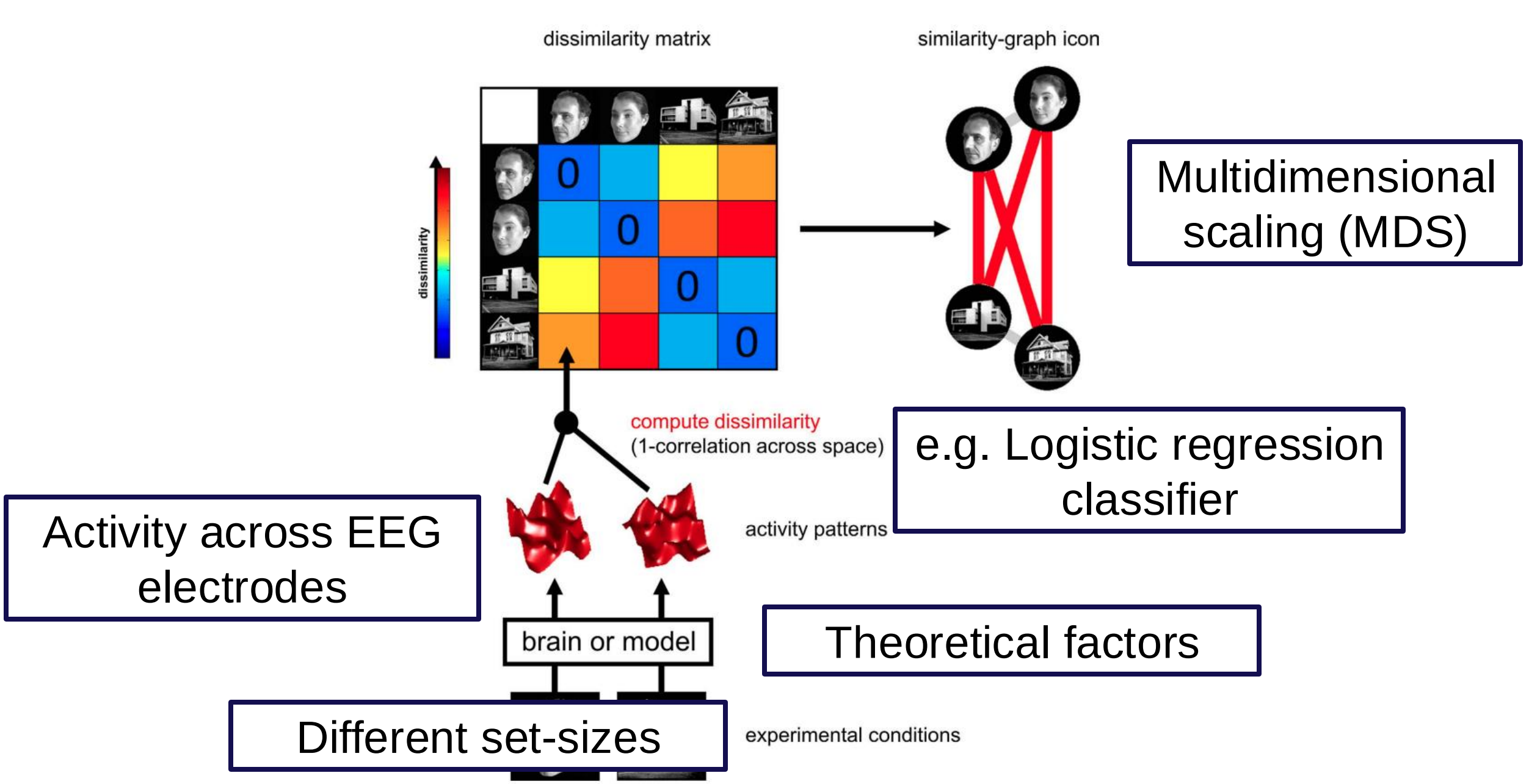
Kriegeskorte, N., Mur, M., & Bandettini, P. A. (2008). Representational similarity analysis-connecting the branches of systems neuroscience. *Frontiers in systems neuroscience*, *2*, 249.

A


computational models
• symbolic models
• connectionist models
• biological neural models
brain-activity data
• cell recordings
• fMRI
• EEG, MEG
representational dissimilarity matrix
behavioral data
• reaction time
• errors
• explicit judgements

B

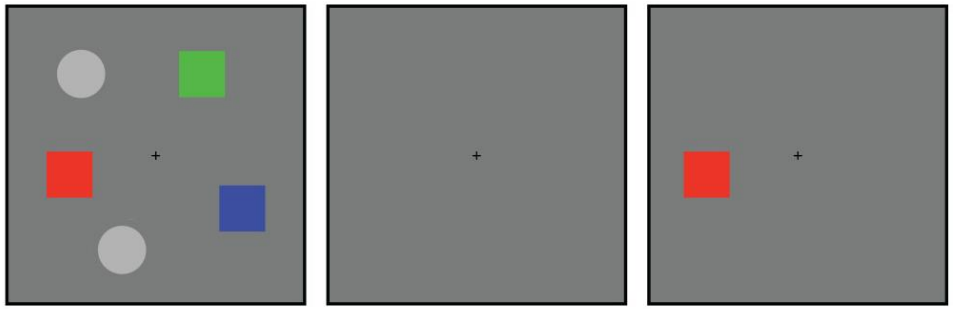

computational model
(stage-2 representation)
voxel-to-unit mapping
unit-to-channel m.
human fMRI
(subject 1, region A)
other modalities
(EEG, MEG, optical imaging etc.)
intersubject mapping
channel-to-neuron m.
human fMRI
(subject 2, region A)
monkey cell recordings
(subject 1, region A)
region-to-region mapping
human fMRI
(subject 2, region B)
monkey fMRI
(subject 1, region B)
voxel-to-neuron mapping
interspecies mapping



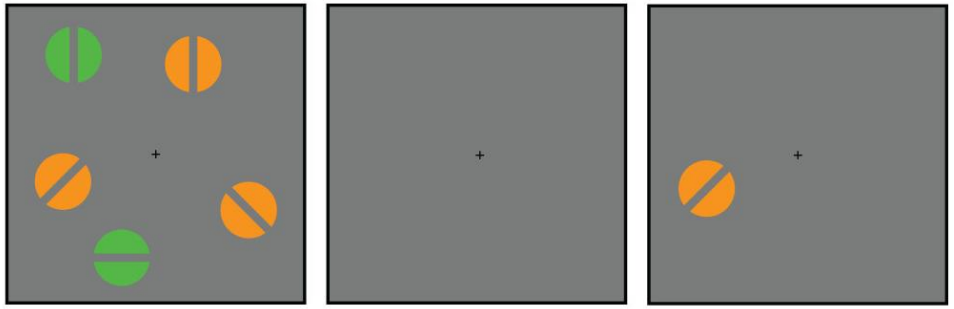
Kriegeskorte, N., Mur, M., & Bandettini, P. A. (2008). Representational similarity analysis-connecting the branches of systems neuroscience. *Frontiers in systems neuroscience*, *2*, 249.

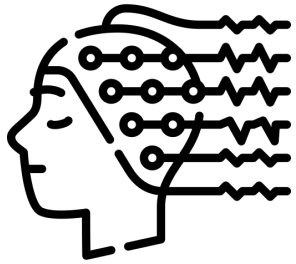
Multivariate classification of working memory

Experiment 1: Color

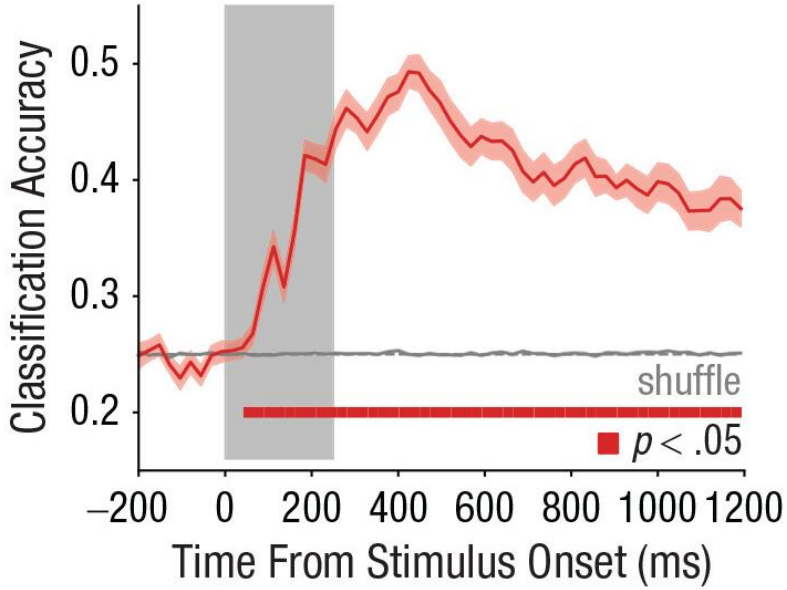


Experiment 2: Orientation

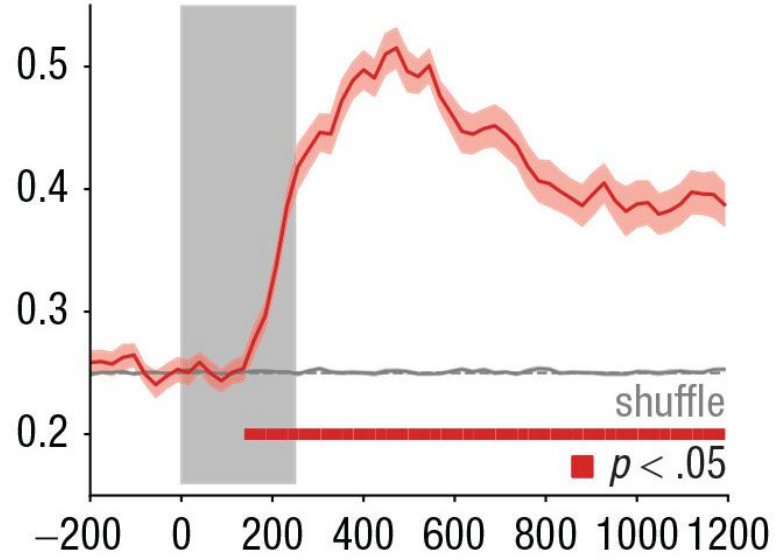




Train and test

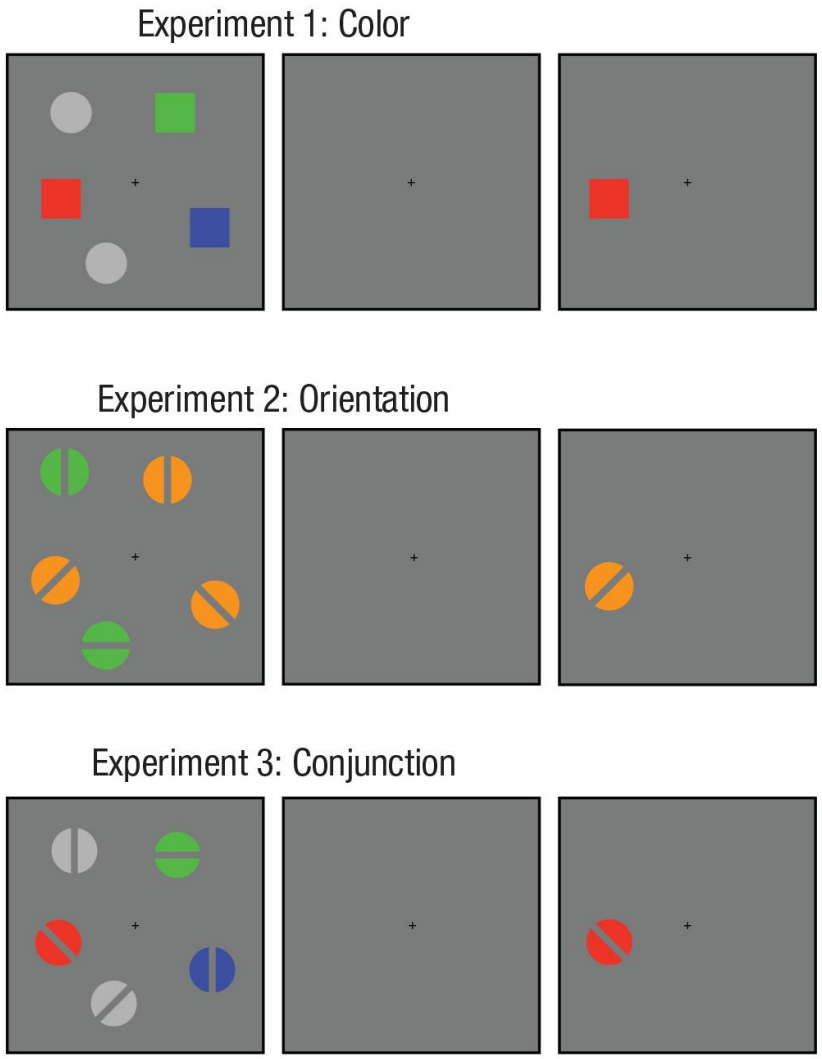


Train and test



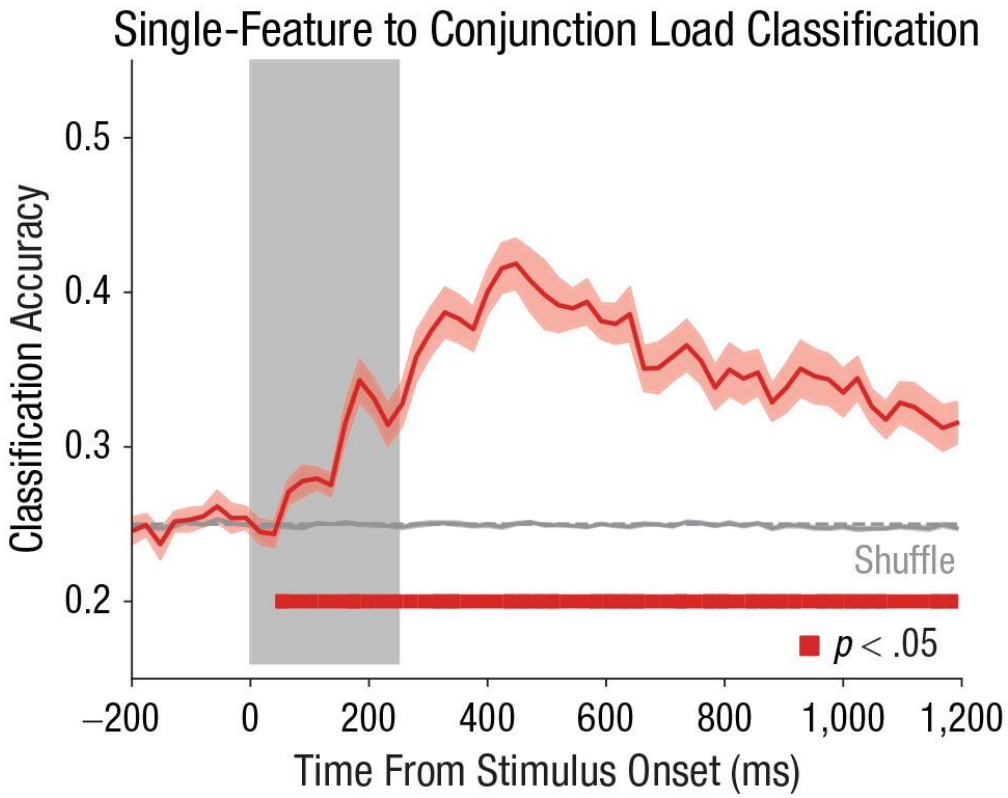
Thyer et al. (2022) *Psychological Science*

Multivariate classification of working memory



Train

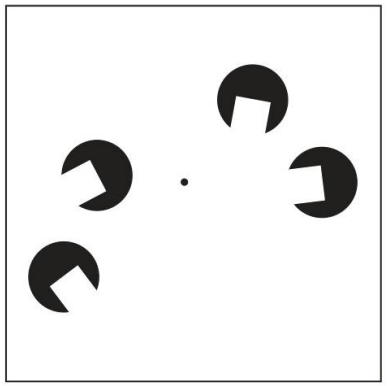
Test



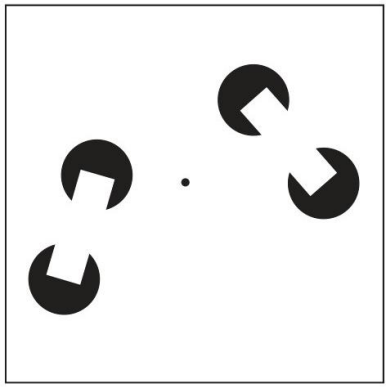
Thyer et al. (2022) *Psychological Science*

Multivariate classification of working memory

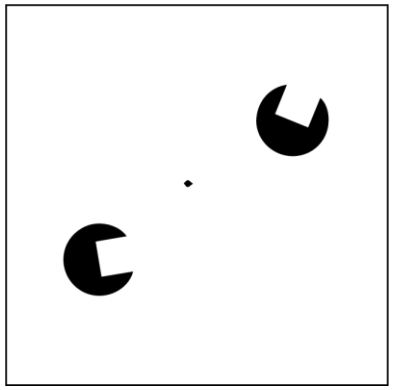
4 Ungrouped

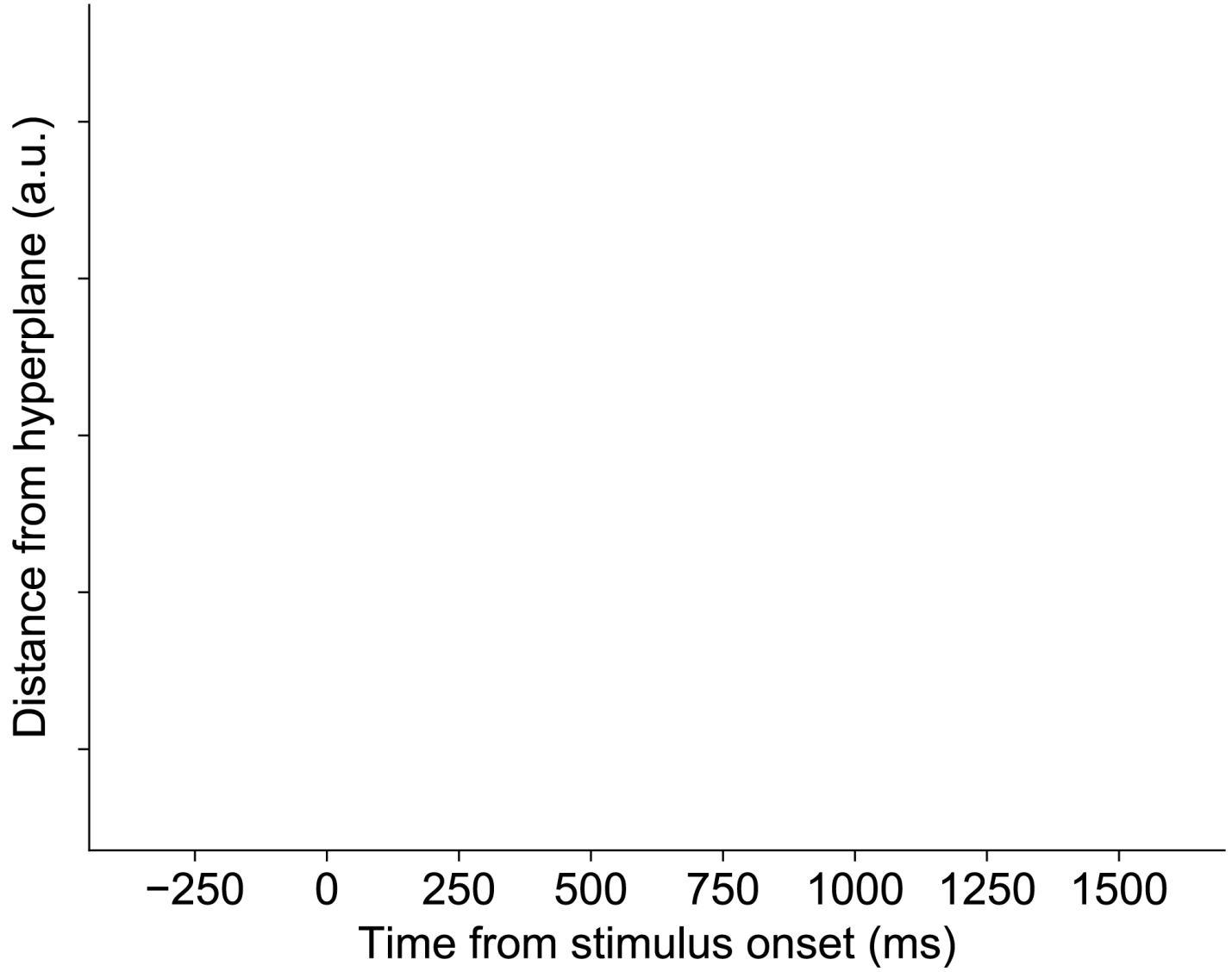


4 Grouped



2 Ungrouped

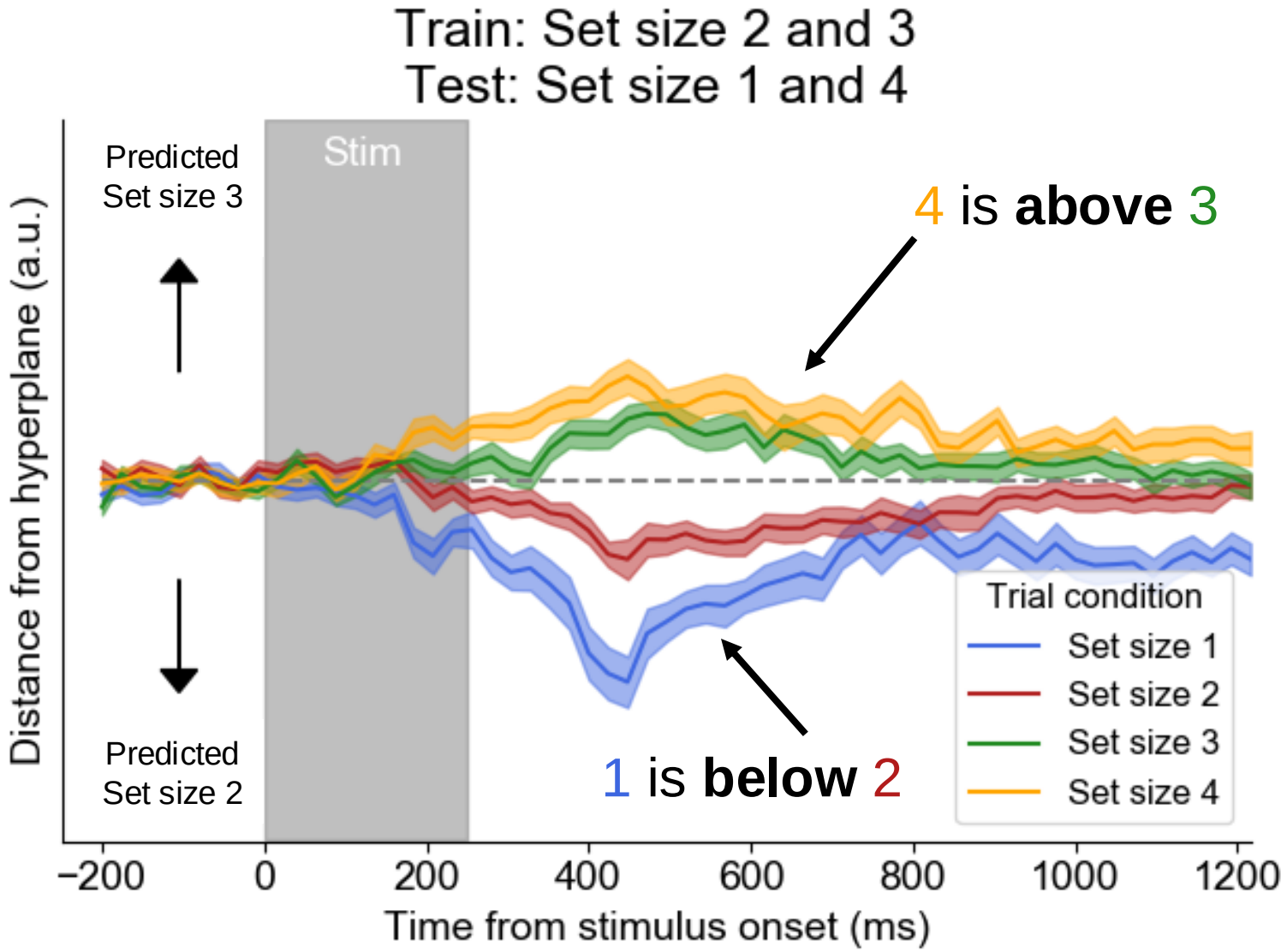




Thyer et al. (2022) *Psychological Science*

Multivariate classification of working memory

Working memory load signal



Does learning change working memory?

Will Epstein

Henry Jones

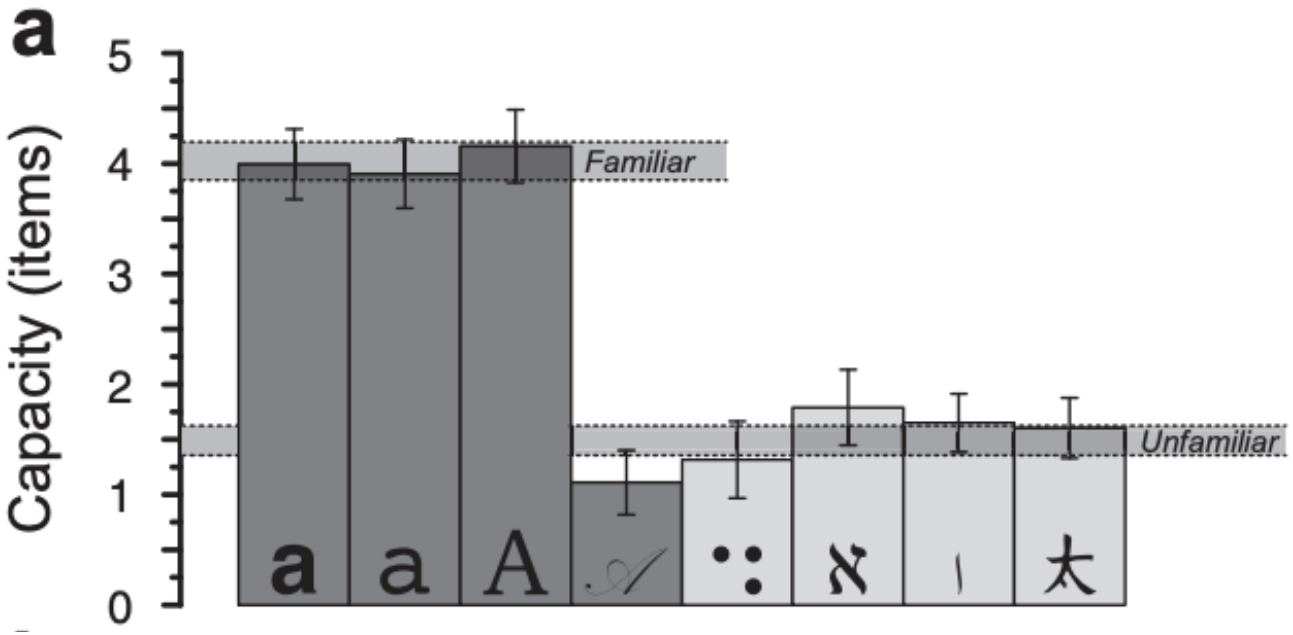
Darius Suplica

William Thyer

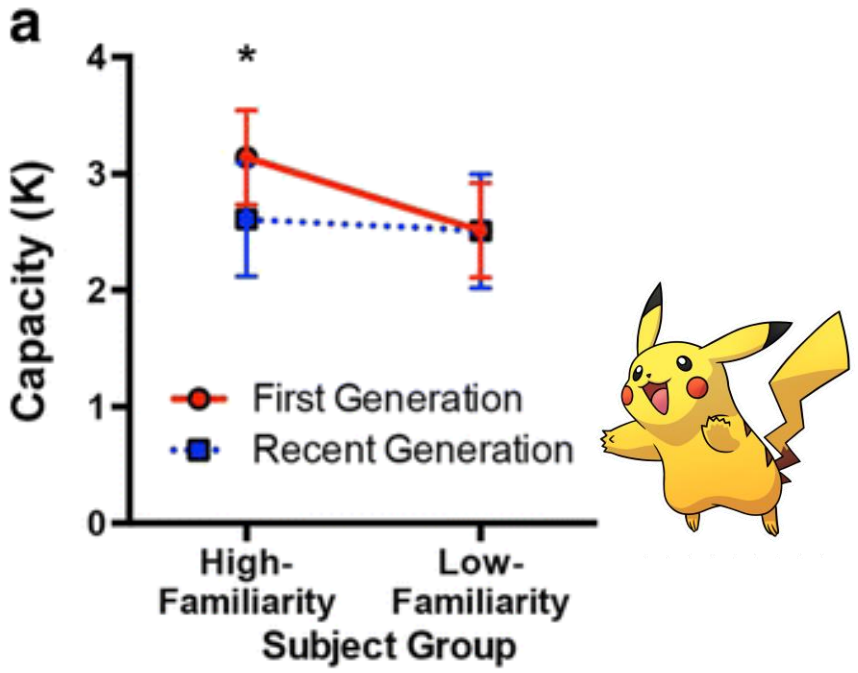
Edward Awh

Working memory is aided by long-term memory

- A hallmark of our visual working memory system is its sharp capacity limit
- But this capacity limit can be overcome with **familiarity:**



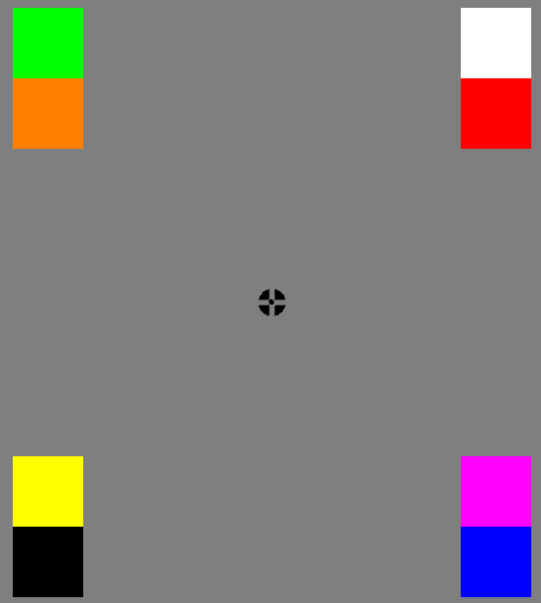
Ngiam et al. (2019) *JEP:G*



Xie and Zhang (2017) *M&C*

Experiment 1: Training

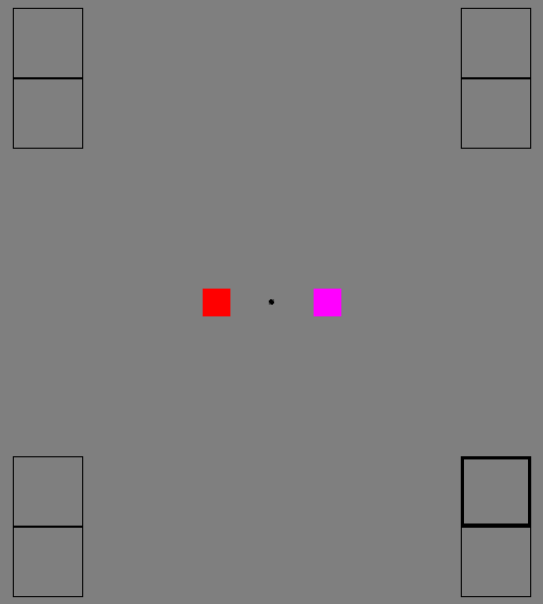
- Subjects completed 600 trials to learn four color pairs:



Experiment 1: Training

Experiment 1: Training

- Two alternative-forced choice – which color was in the bolded location?

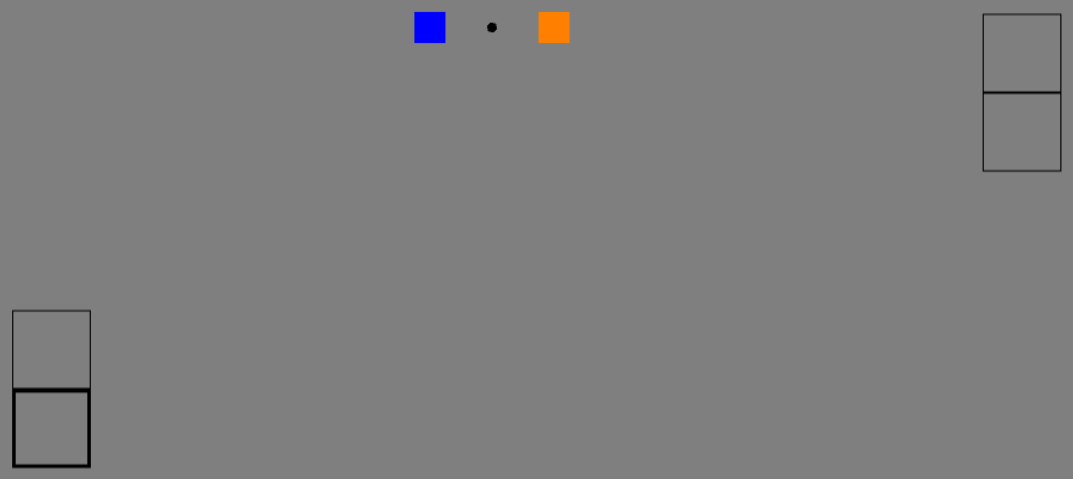


Experiment 1: Pre-training and post-training

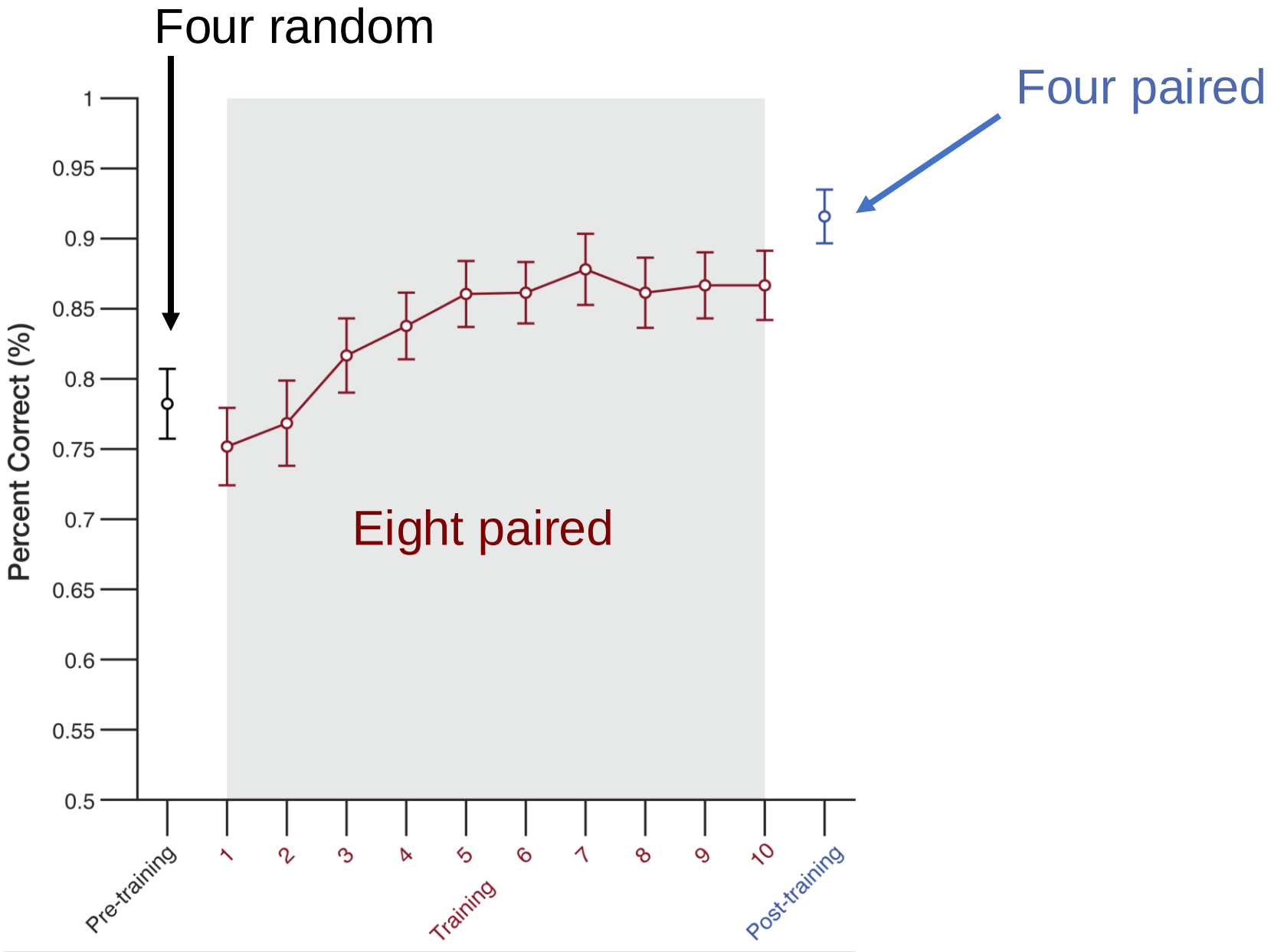
- Before training – 4 random colors
- After training – 4 paired colors (two learned pairs)

Experiment 1: Pre-training and post-training

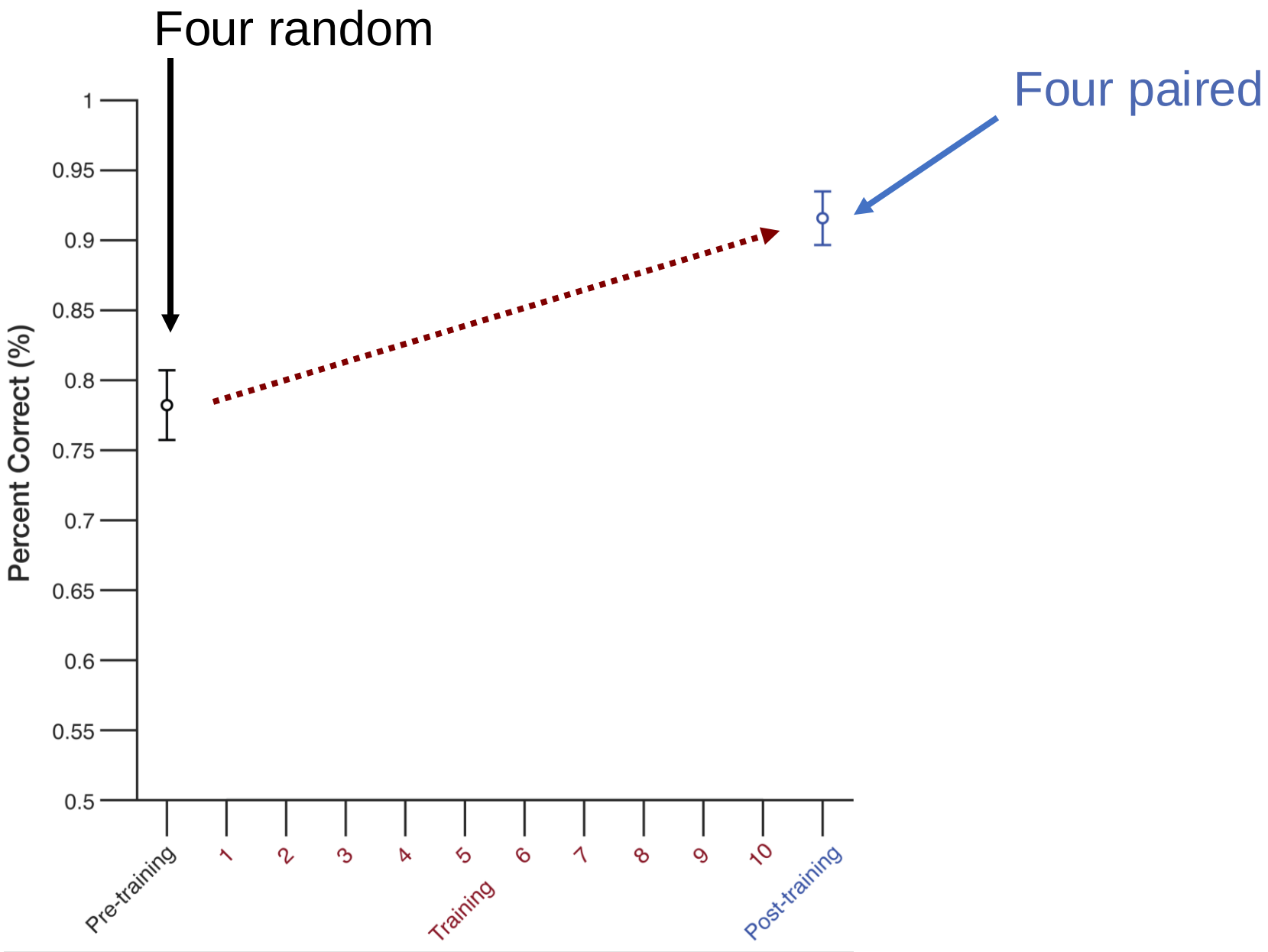
Experiment 1: Pre-training and post-training



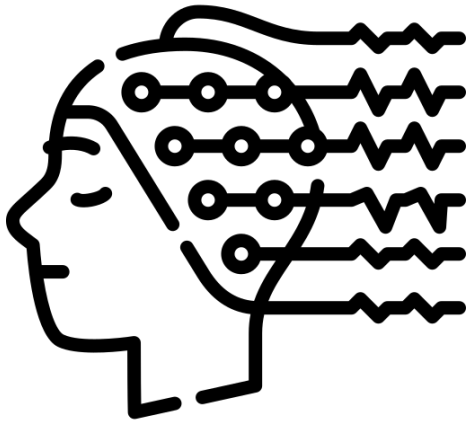
E1: training session – aggregate performance

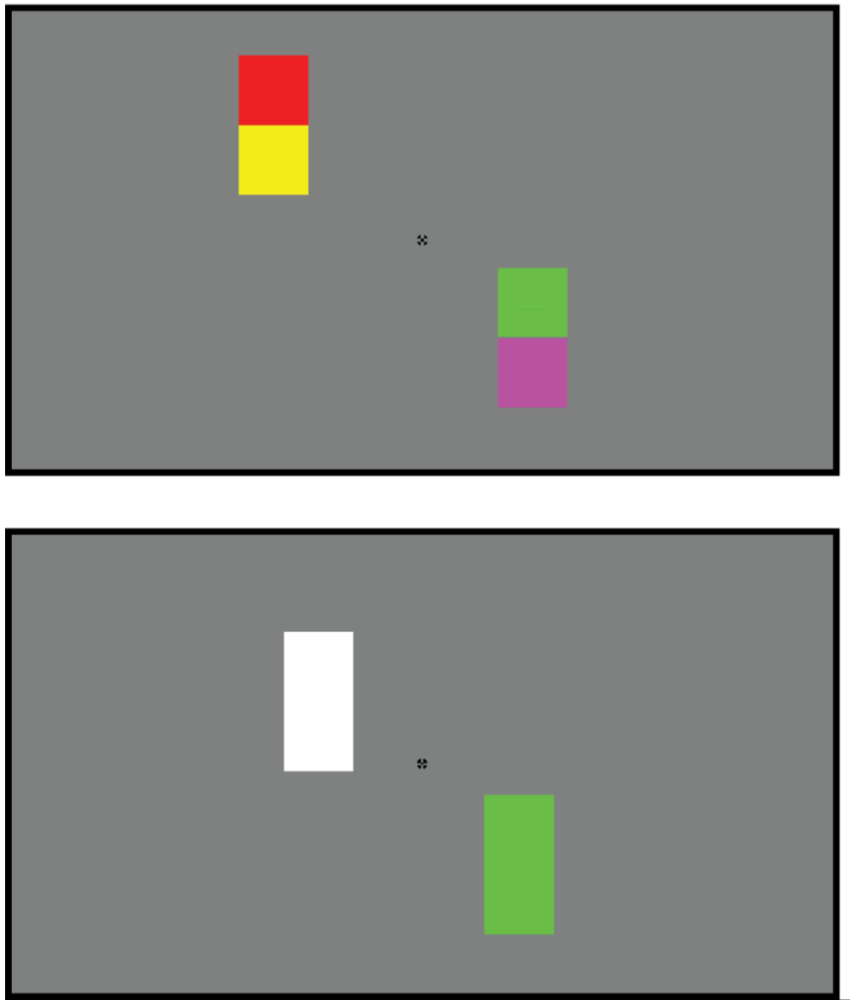


E1: training session – average performance



Experiment 1: EEG session





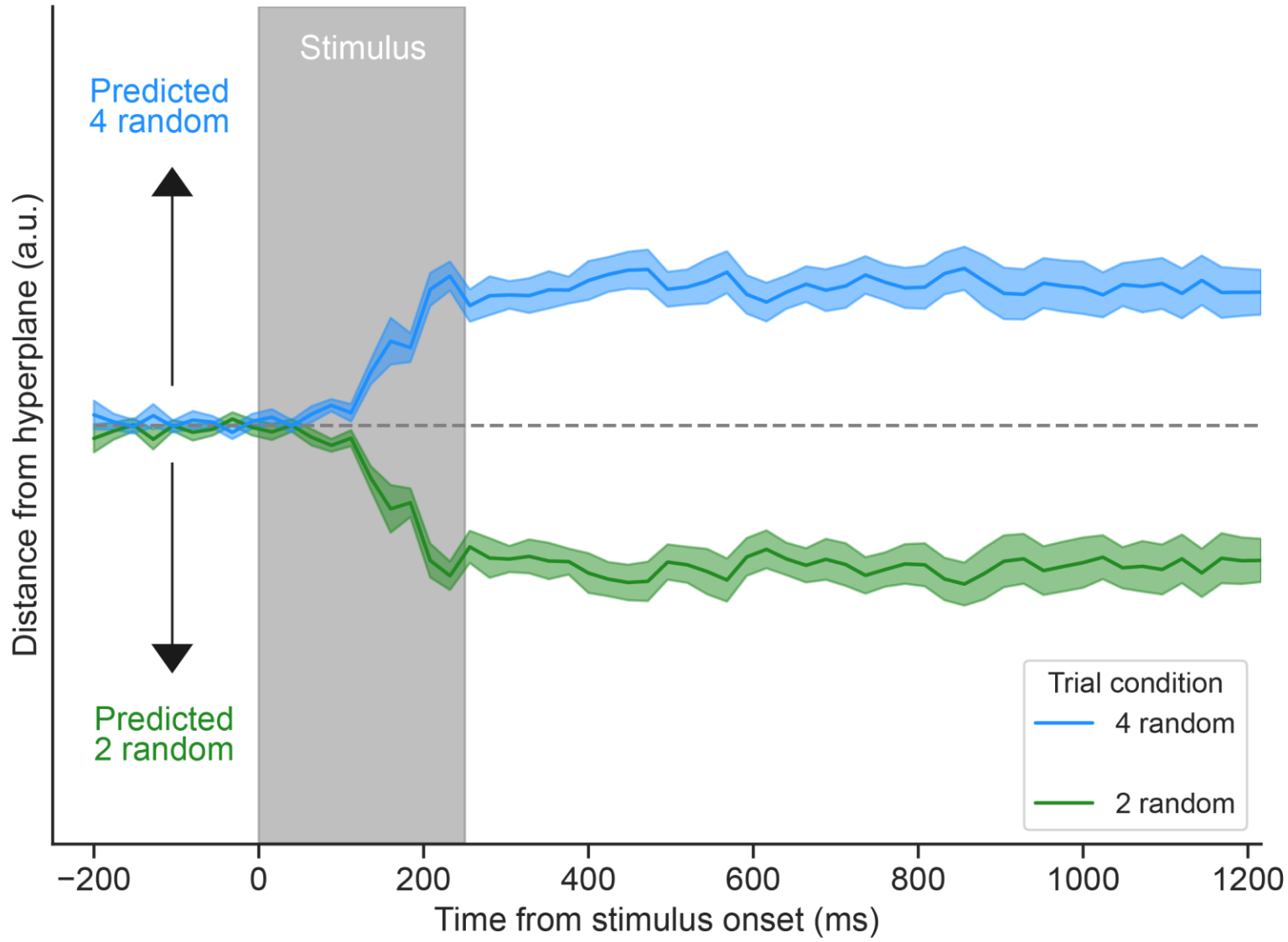
Four random

Four paired

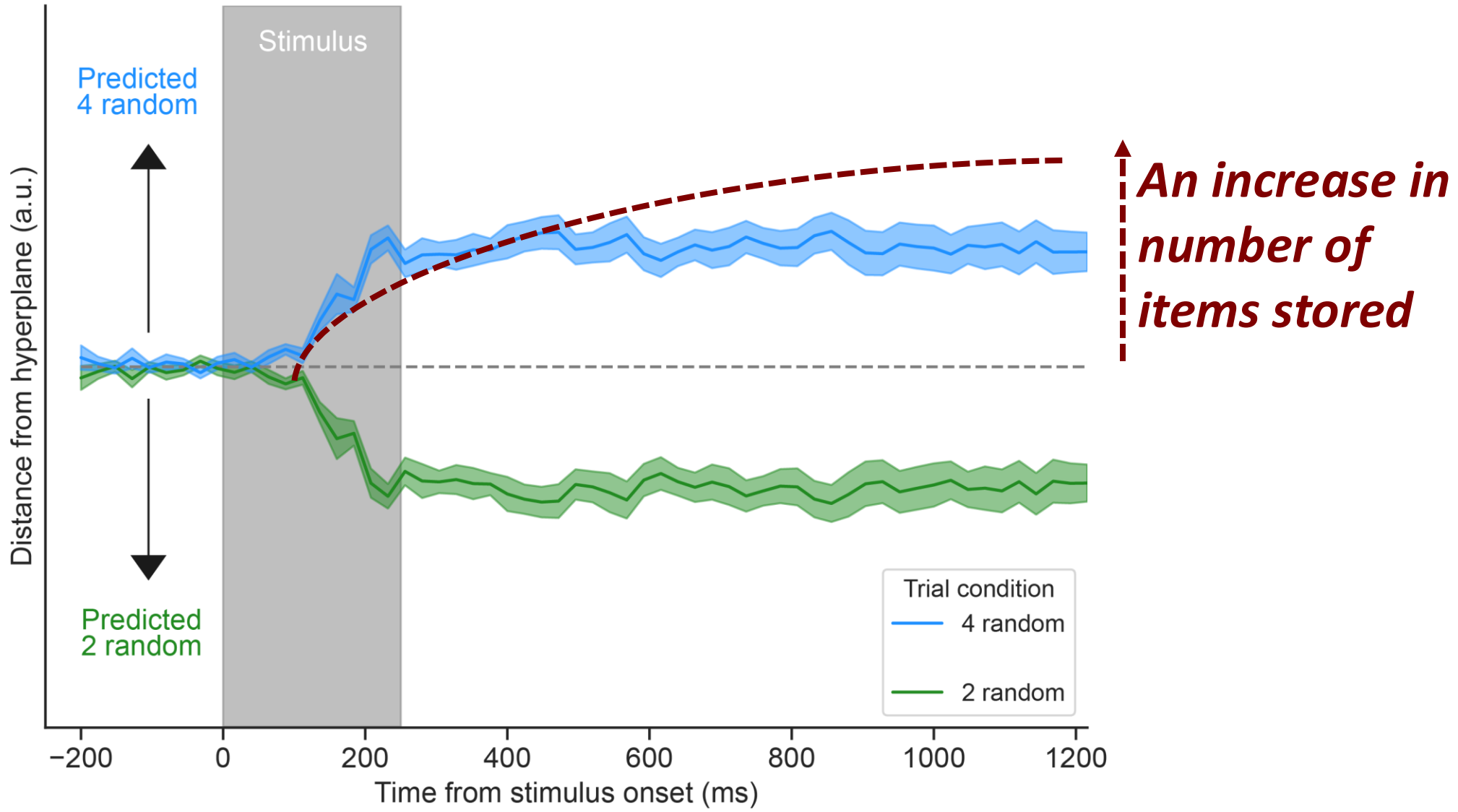
Two random

Perceptually equivalent

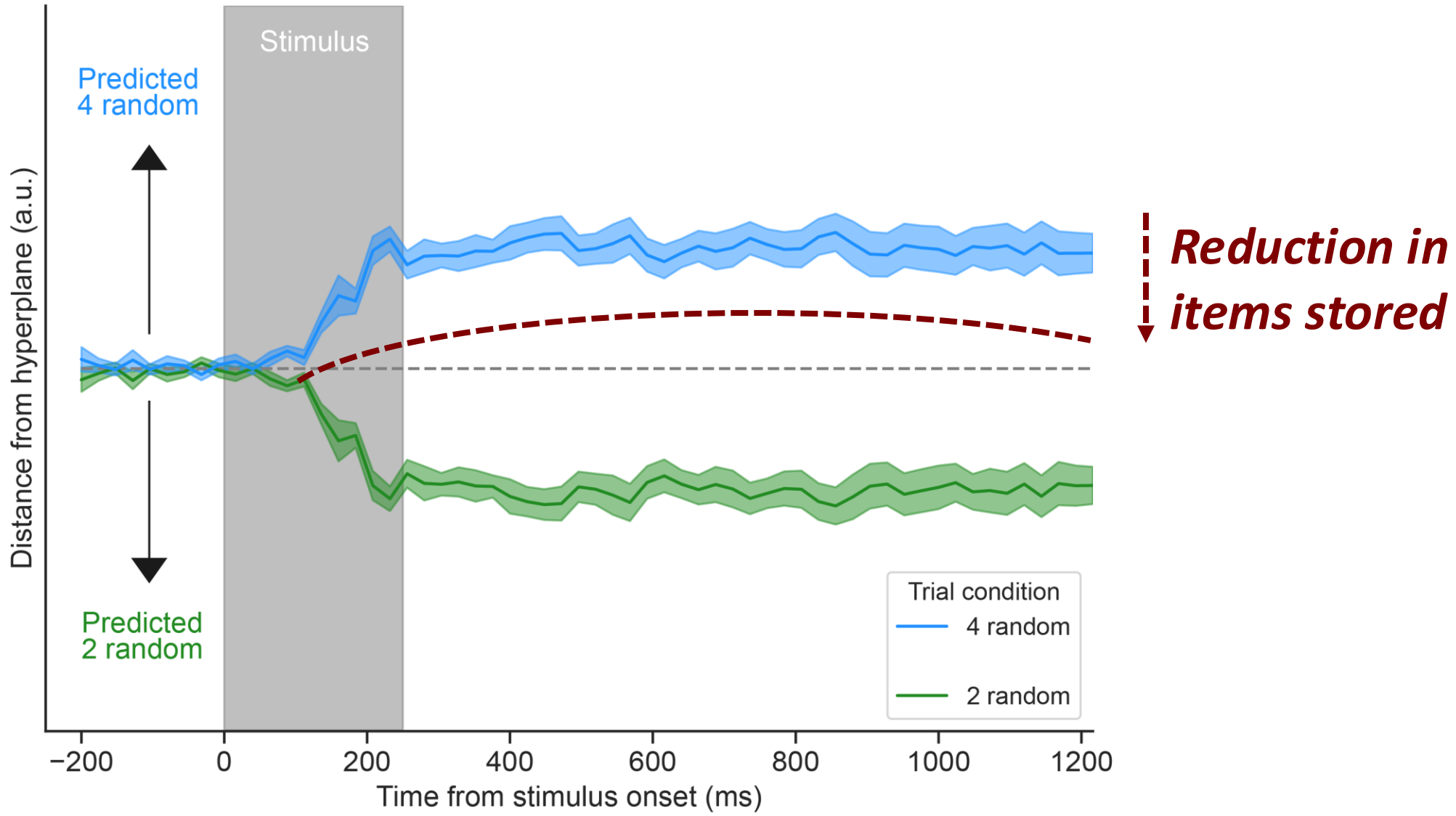
Train 2 random versus 4 random



Test on 4 paired?

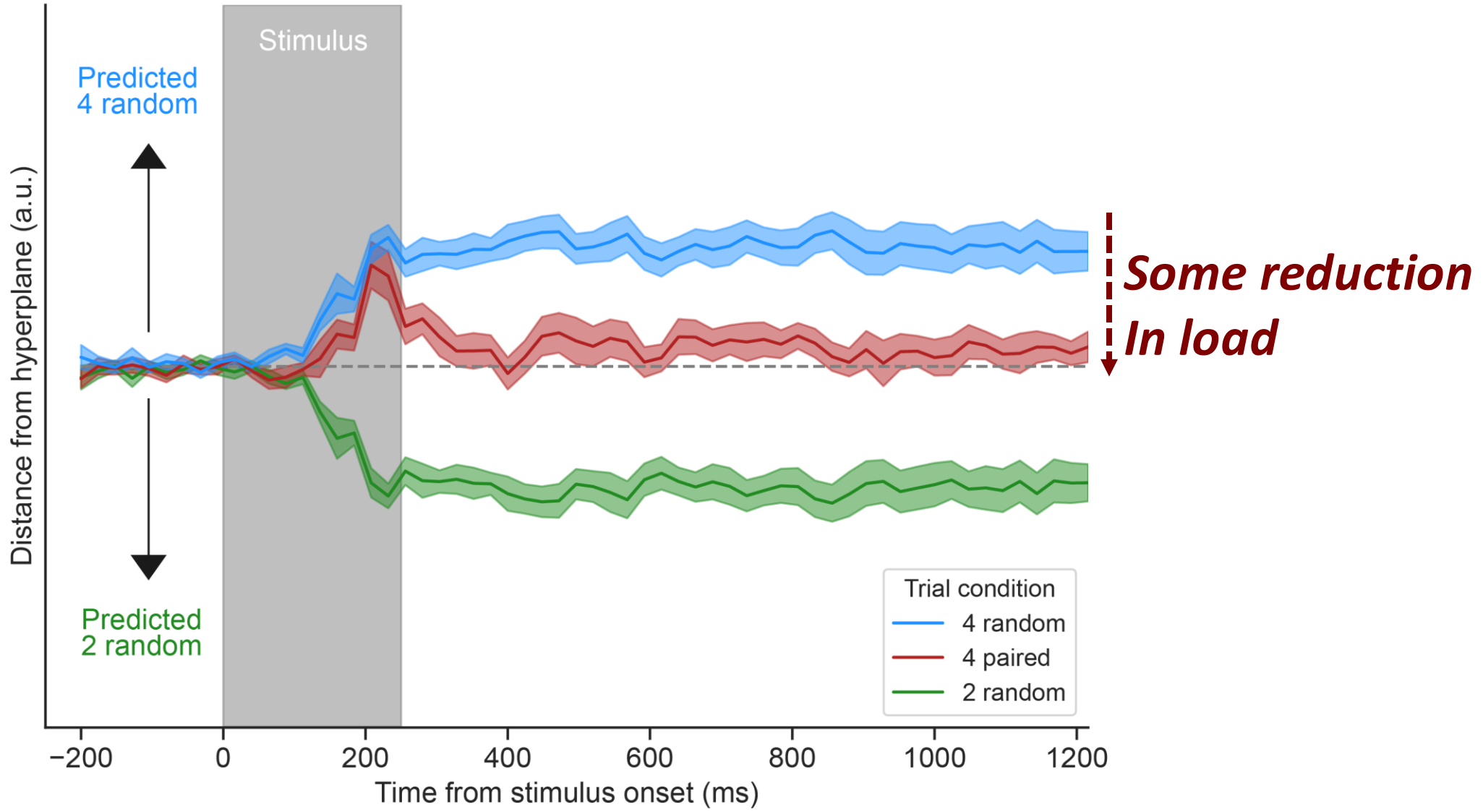


Test on 4 paired?

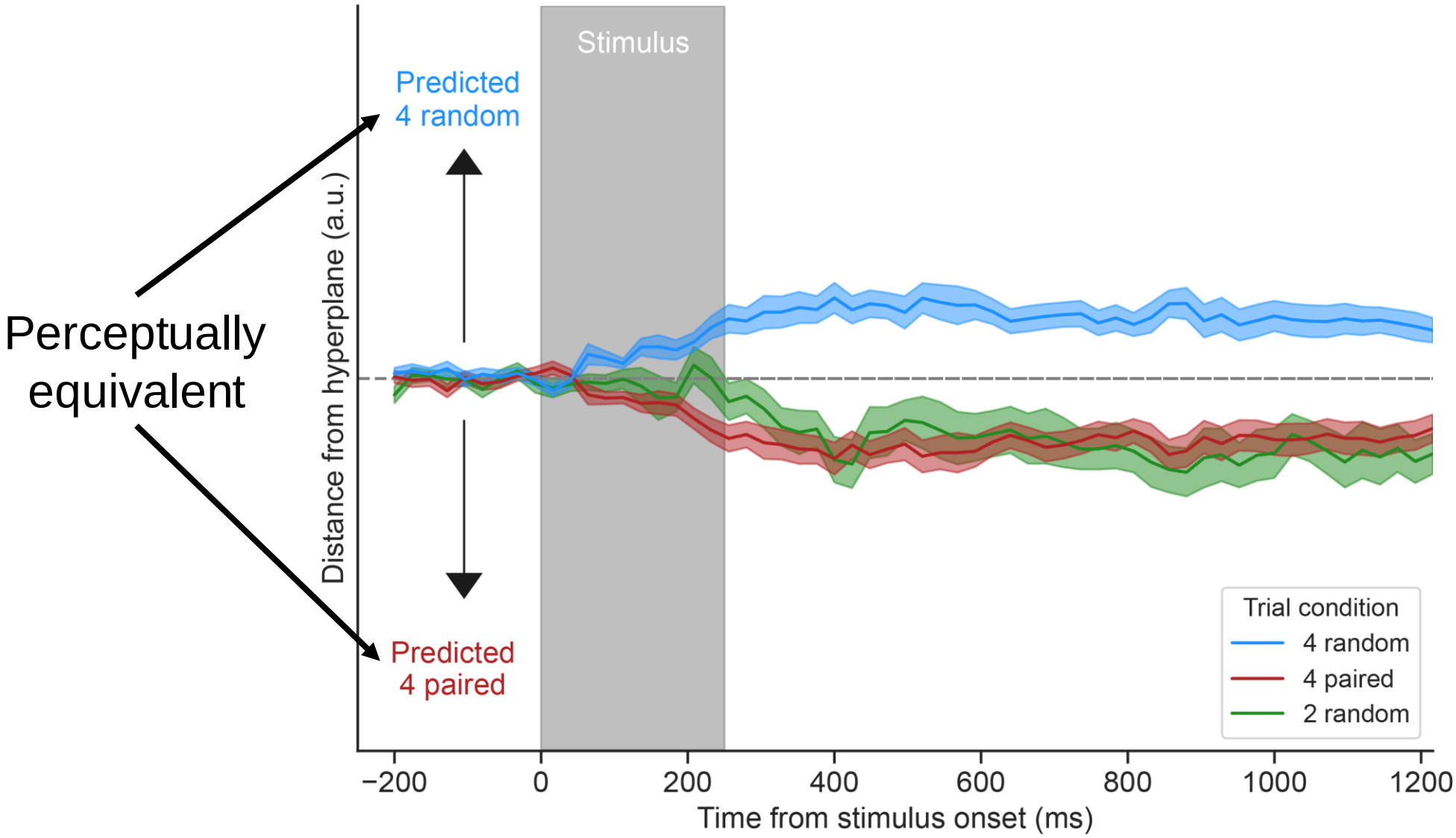


Reduction in items stored

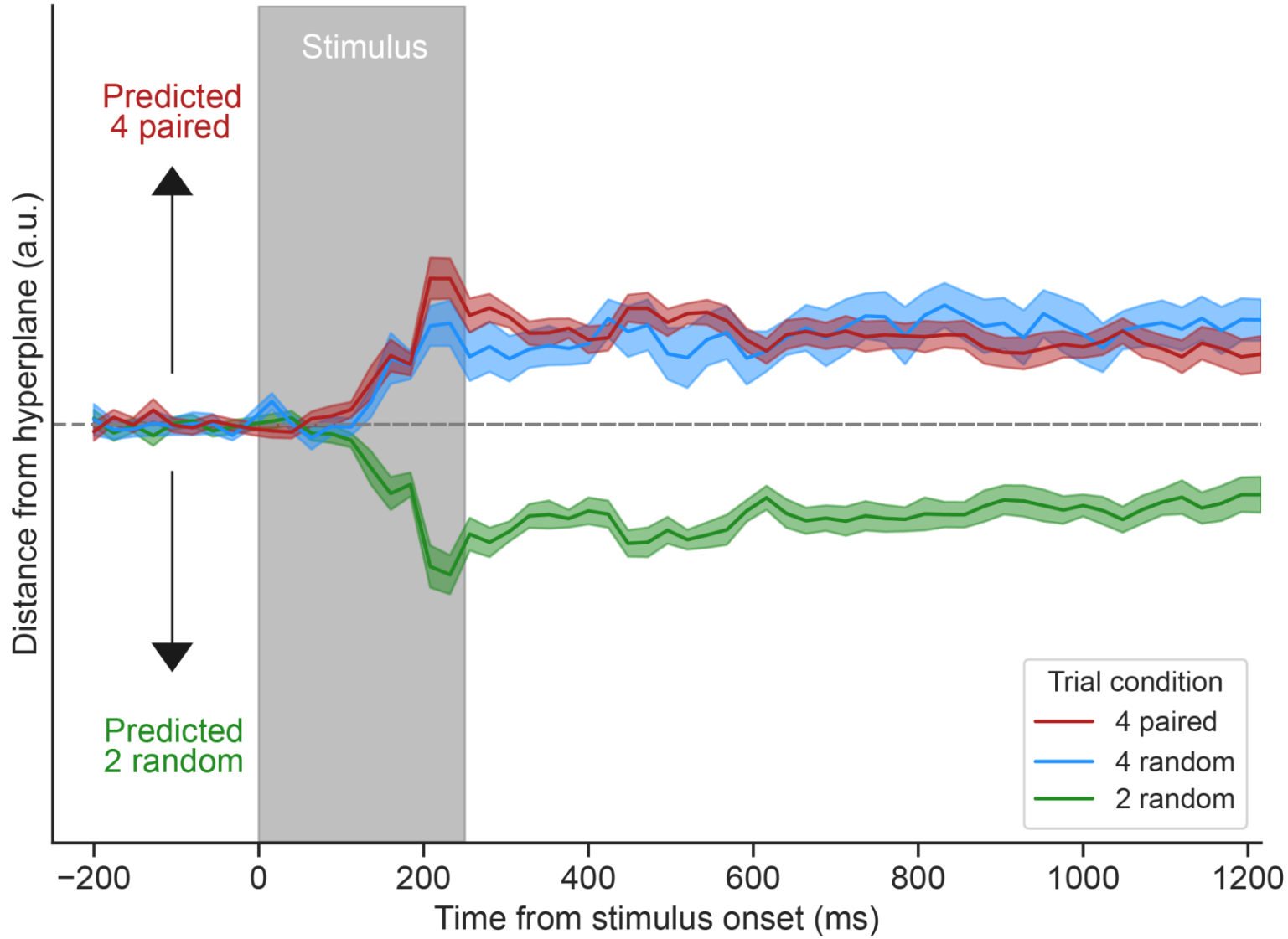
Train 2 random versus 4 random, test 4 paired



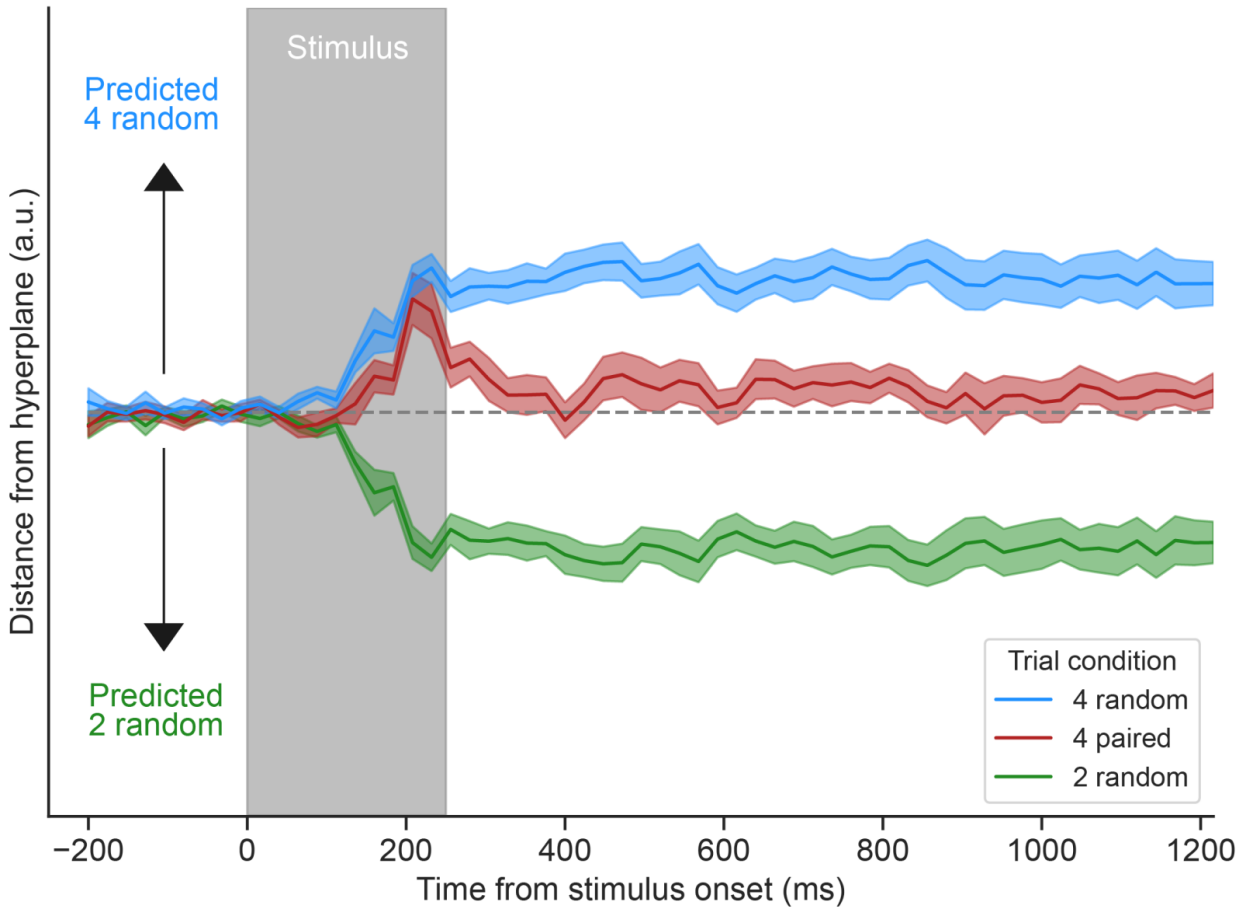
Train 4 random versus 4 paired, test 2 random

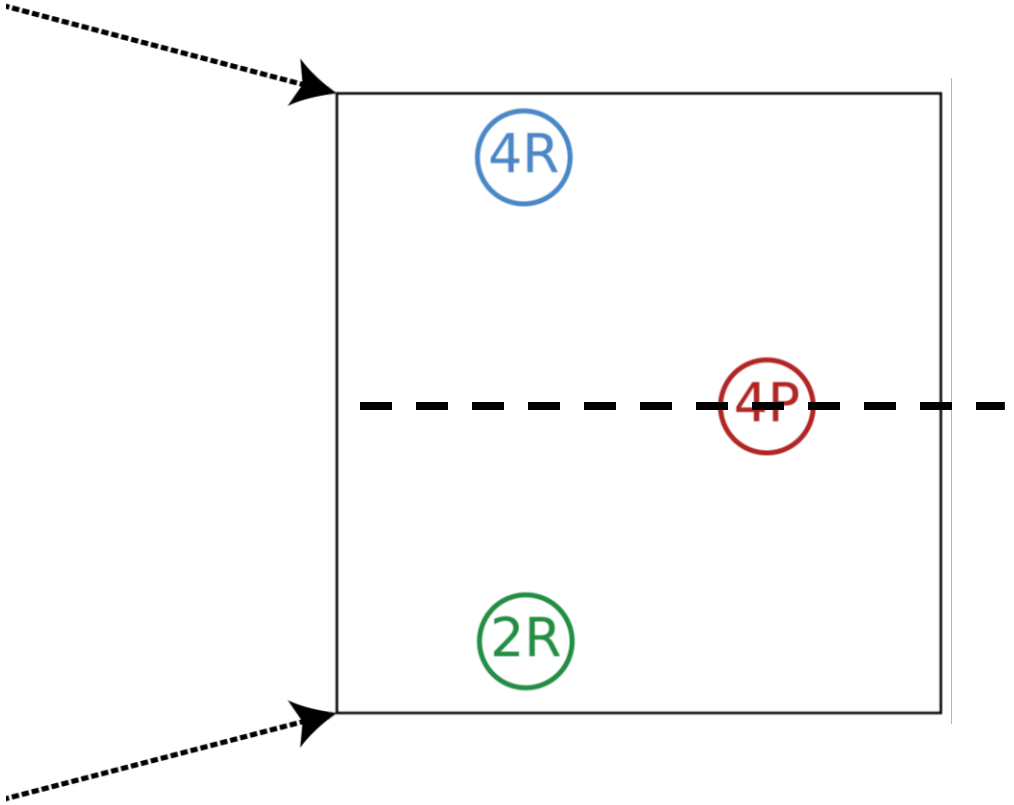


Train 2 random versus 4 paired, test 4 random

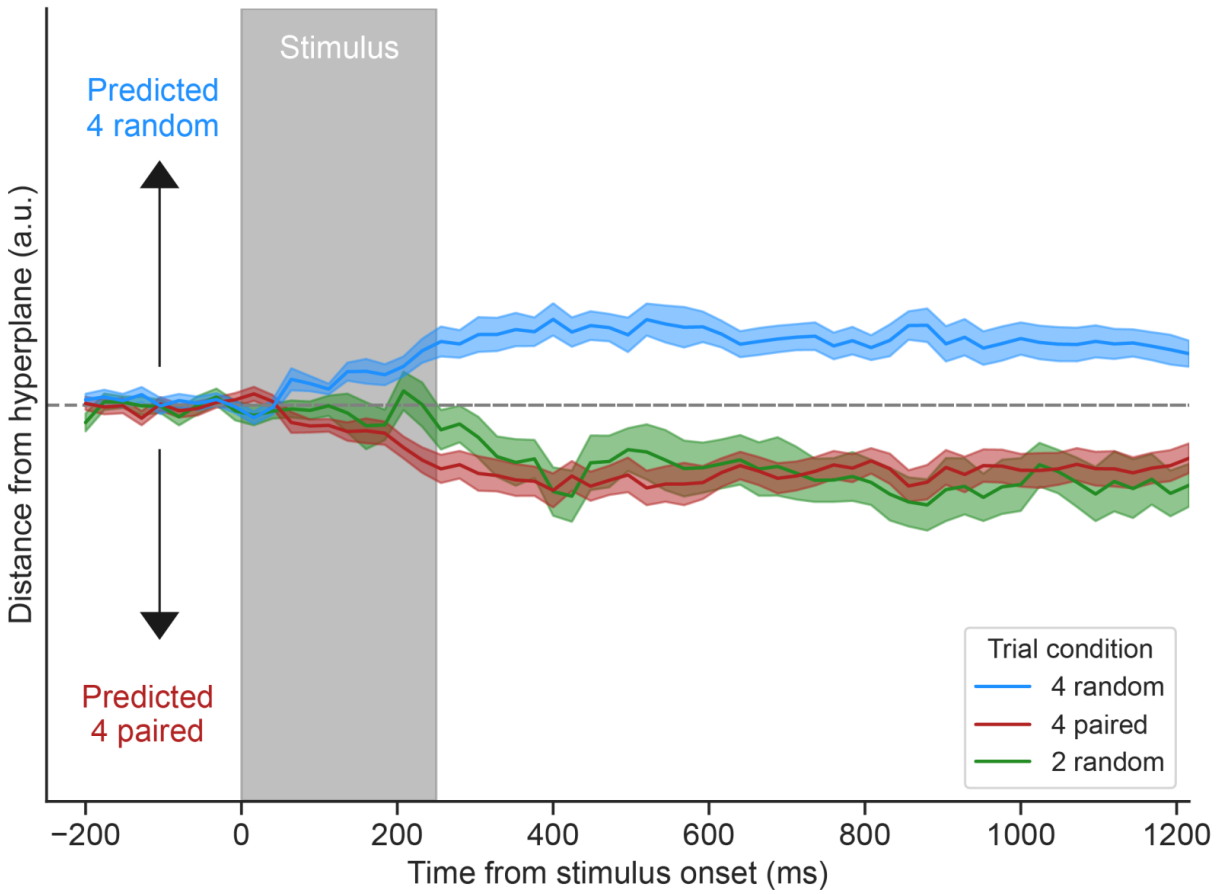


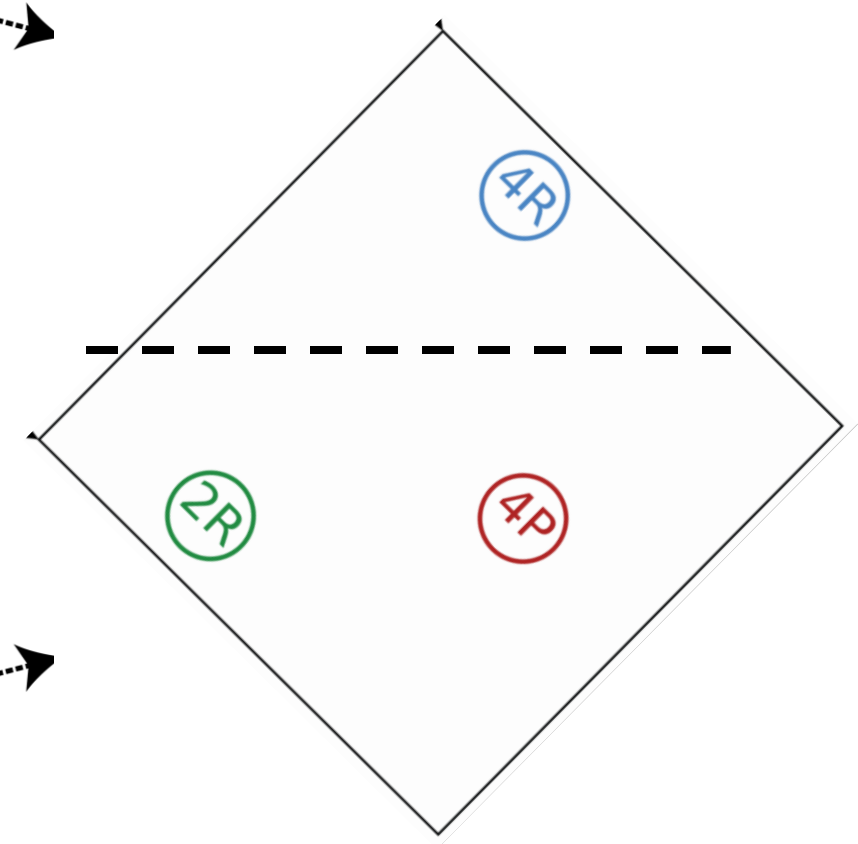
Multidimensional scaling



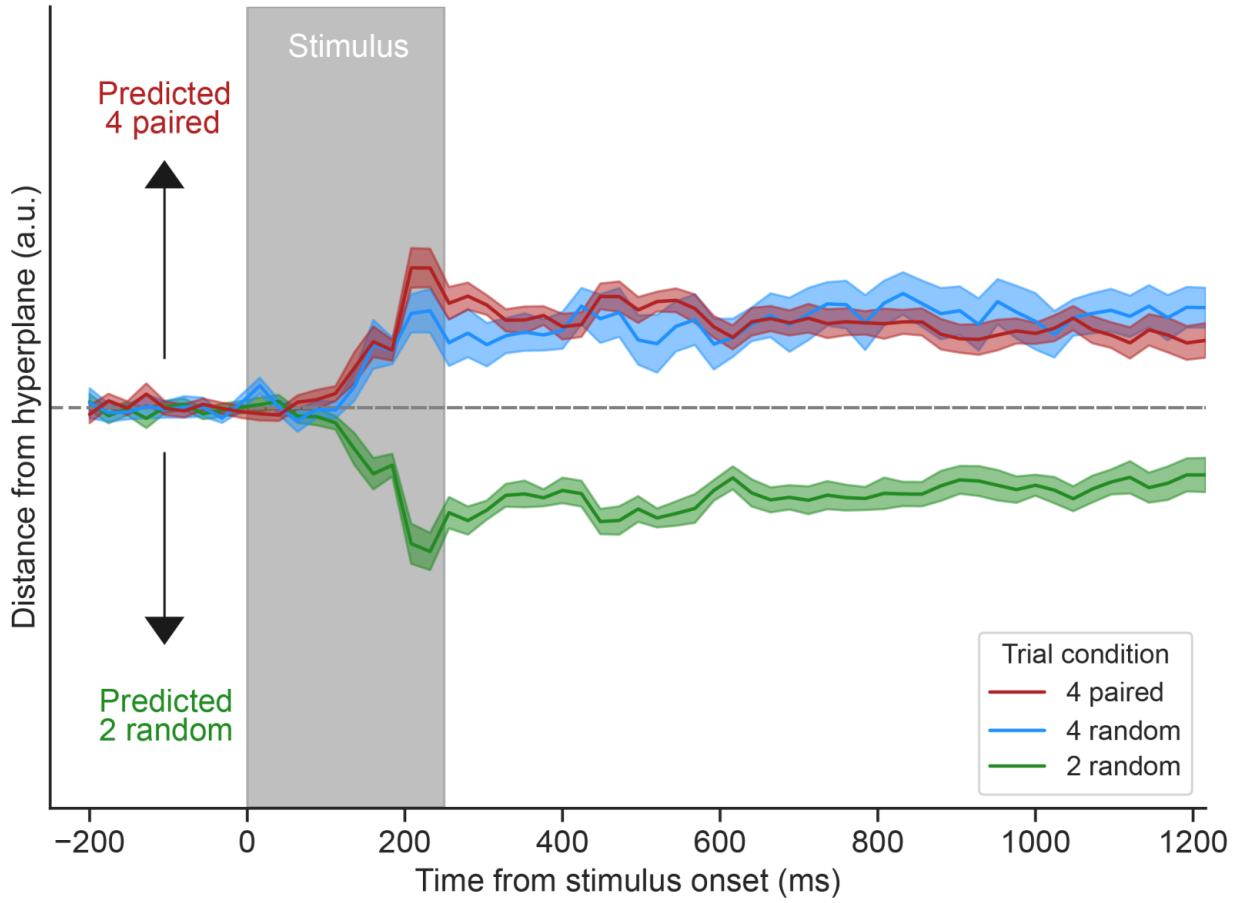


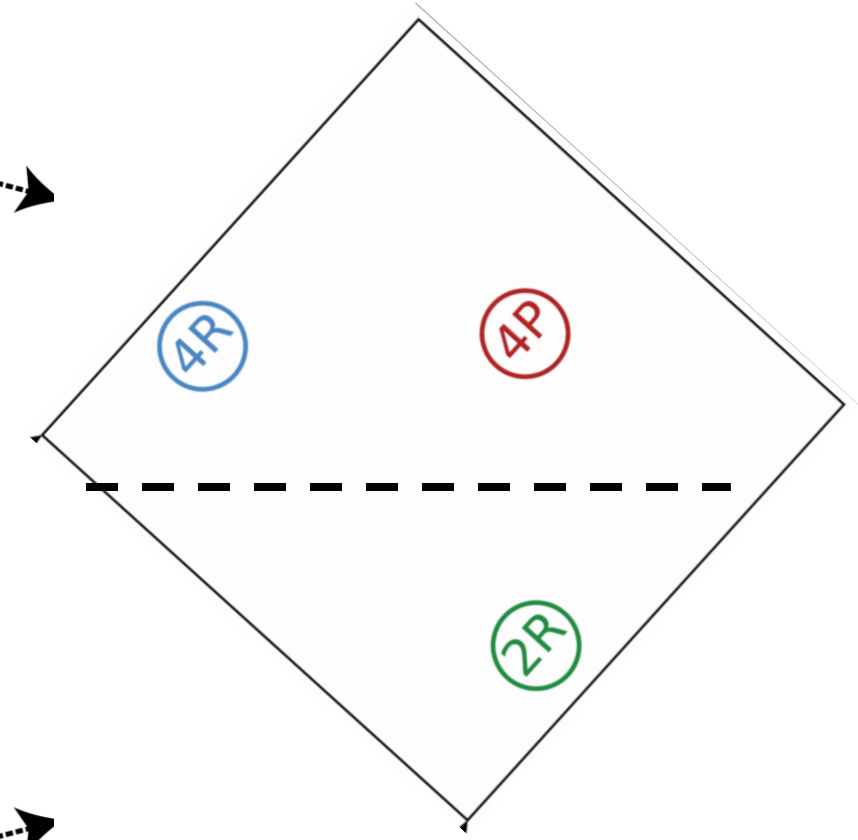
Multidimensional scaling





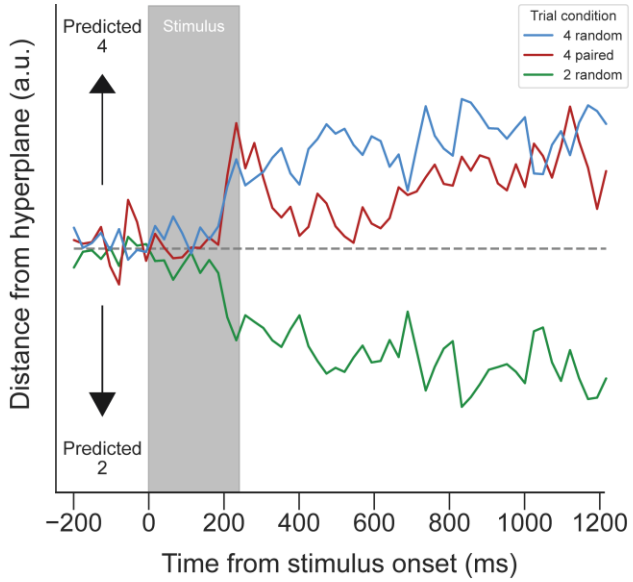
Multidimensional scaling

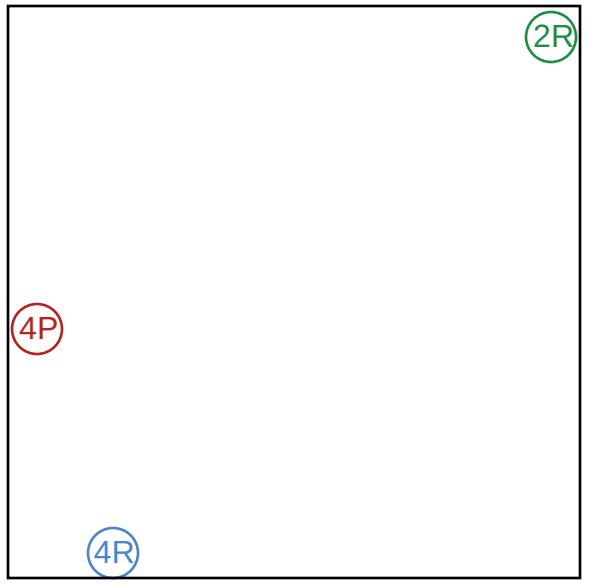




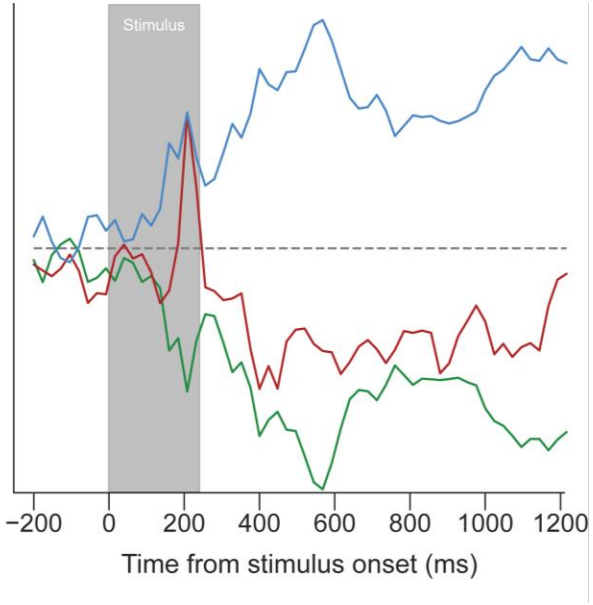
Individual differences

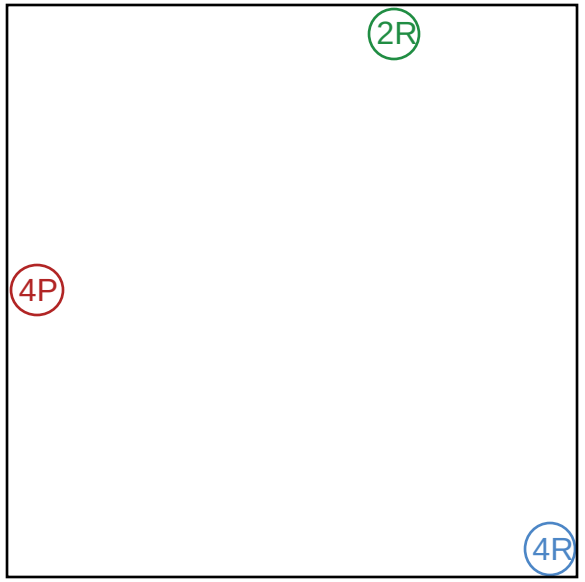
"Weak chunking"





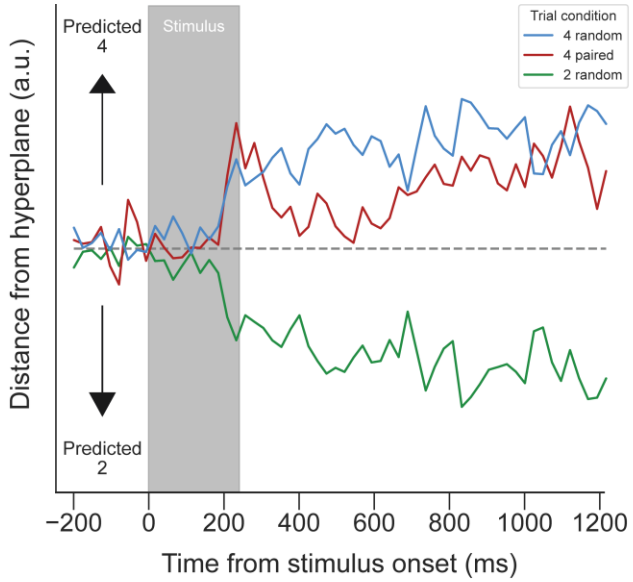
"Strong chunking"

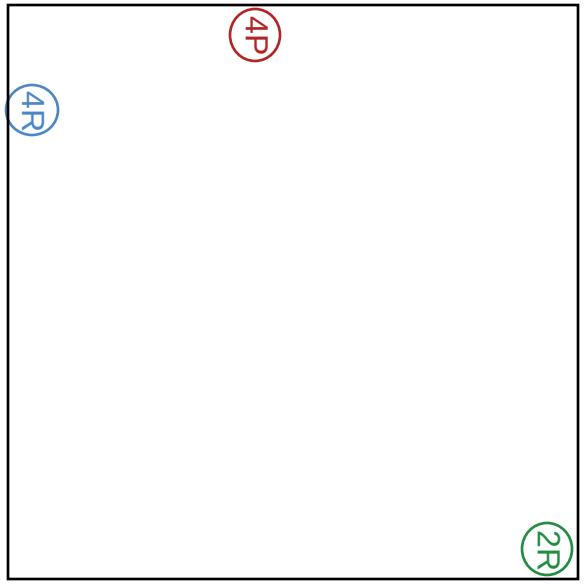




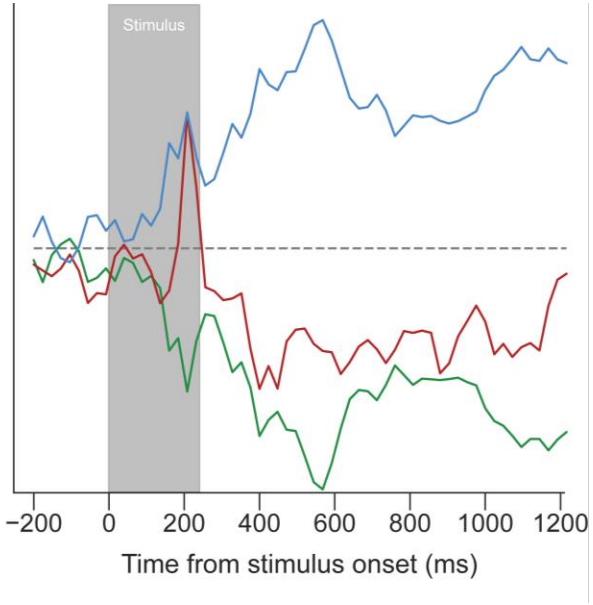
Individual differences

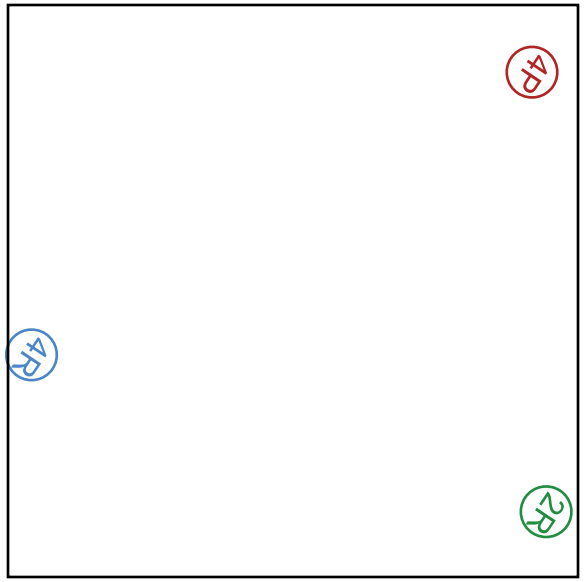
"Weak chunking"





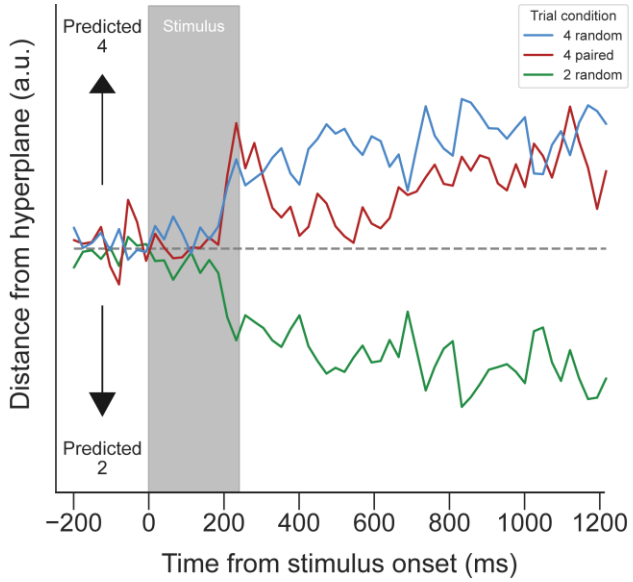
"Strong chunking"

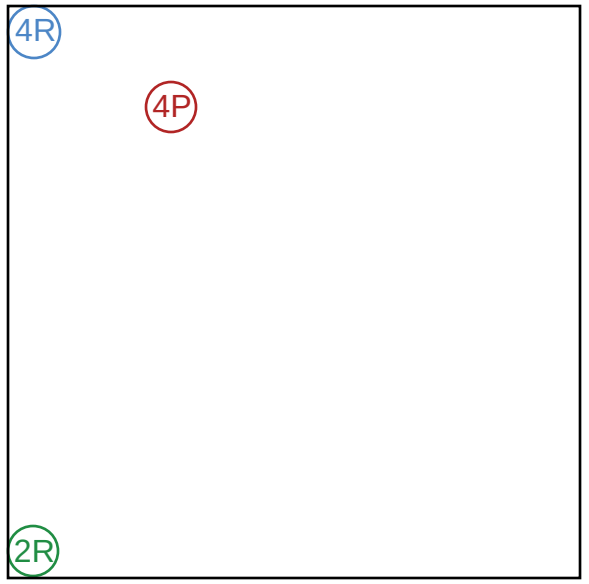




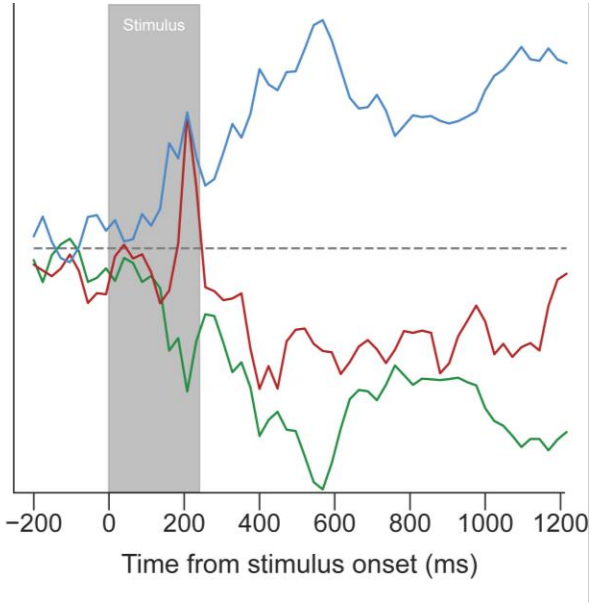
Individual differences

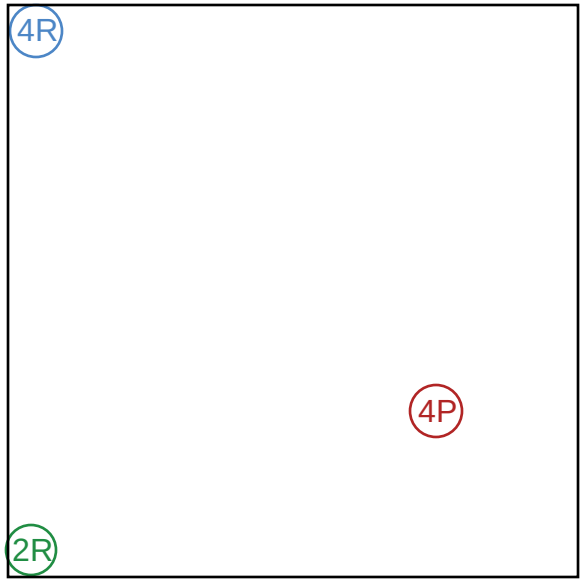
"Weak chunking"





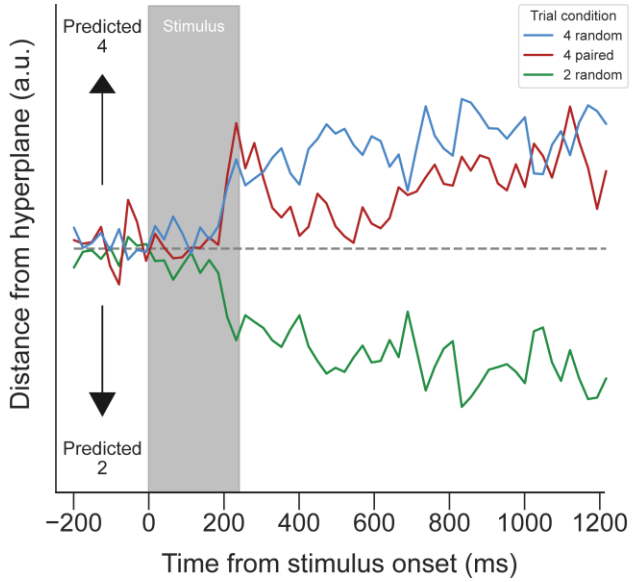
"Strong chunking"



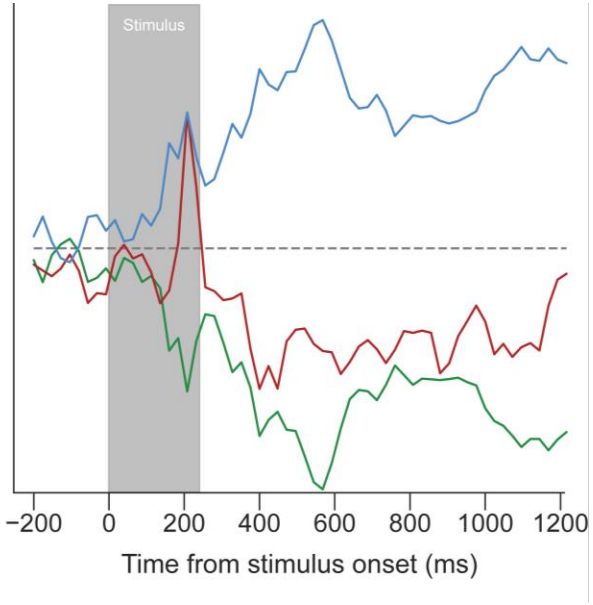


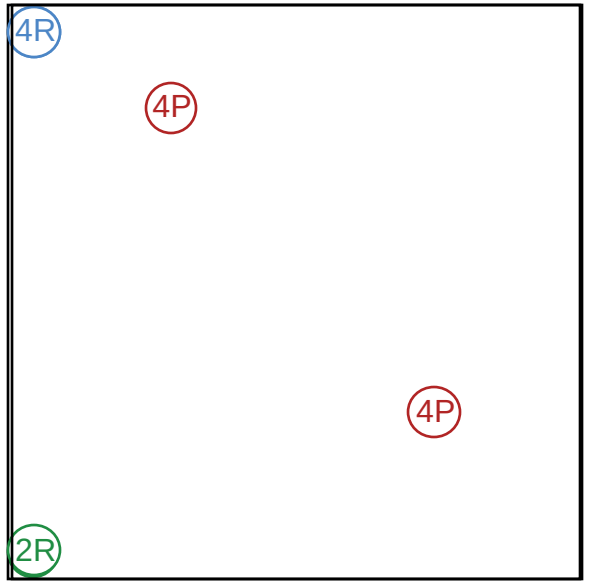
Individual differences

"Weak chunking"

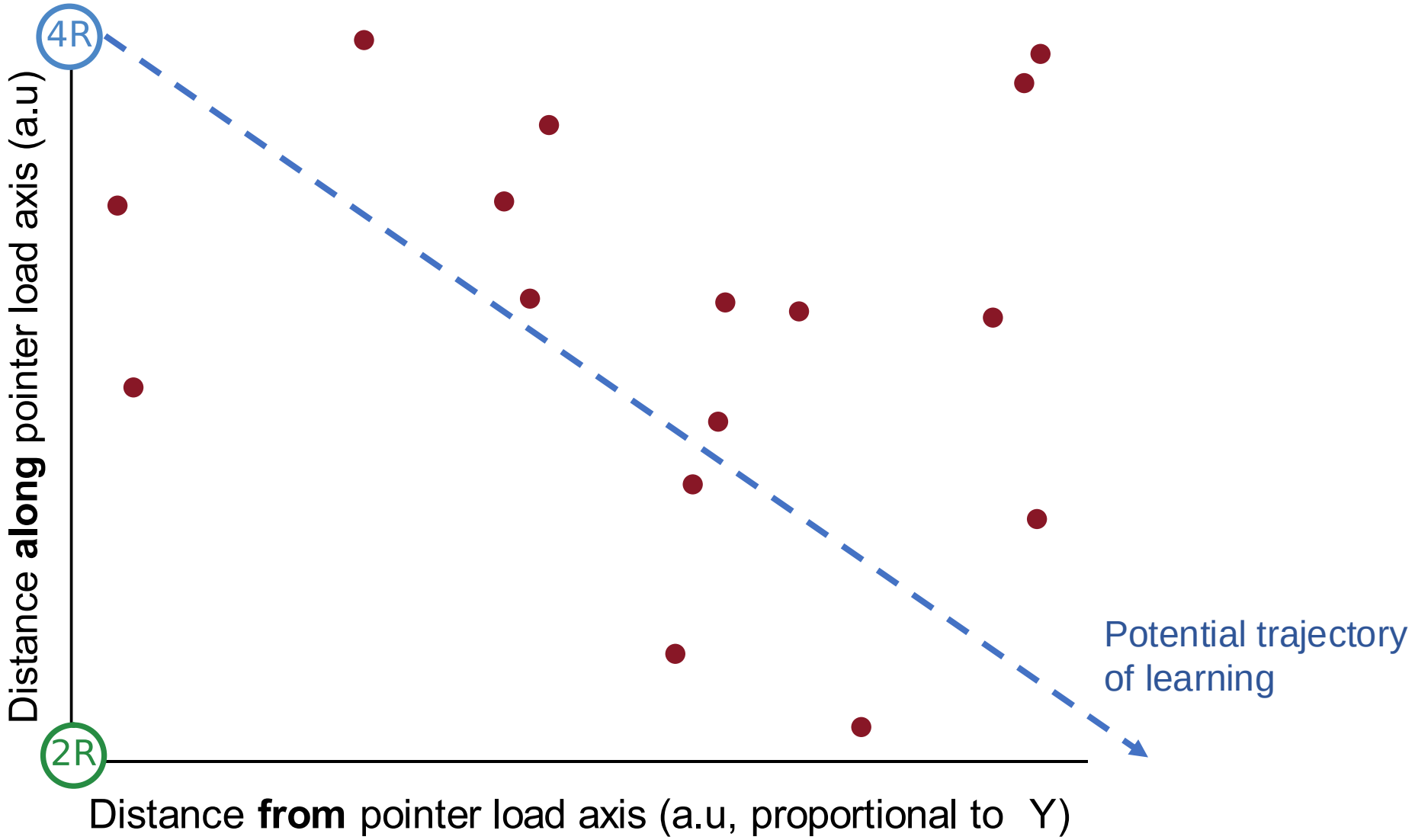


"Strong chunking"



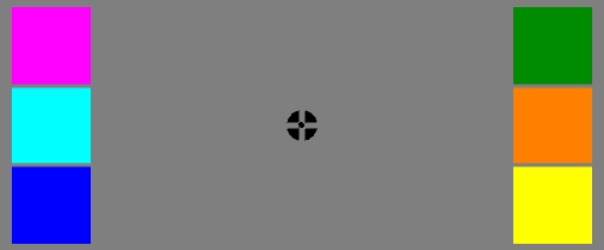


Individual differences



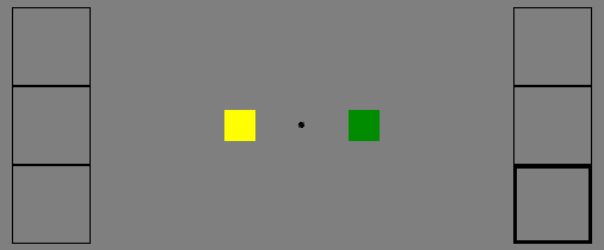
Training

- Trained subjects to learn three color triplets

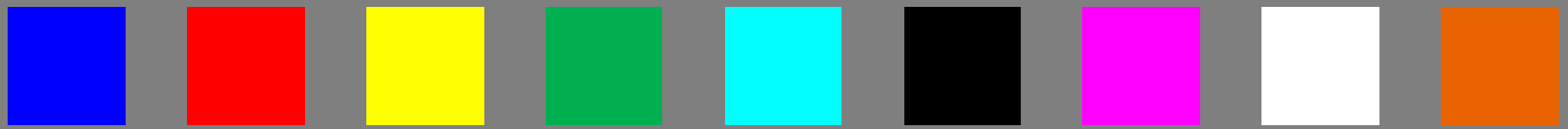


Training

Training

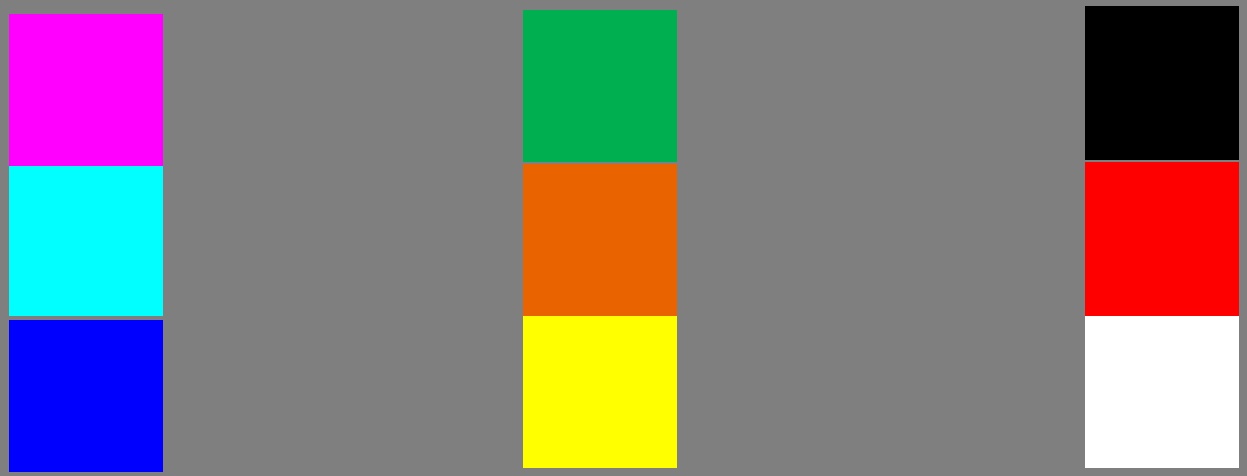


Awareness Test

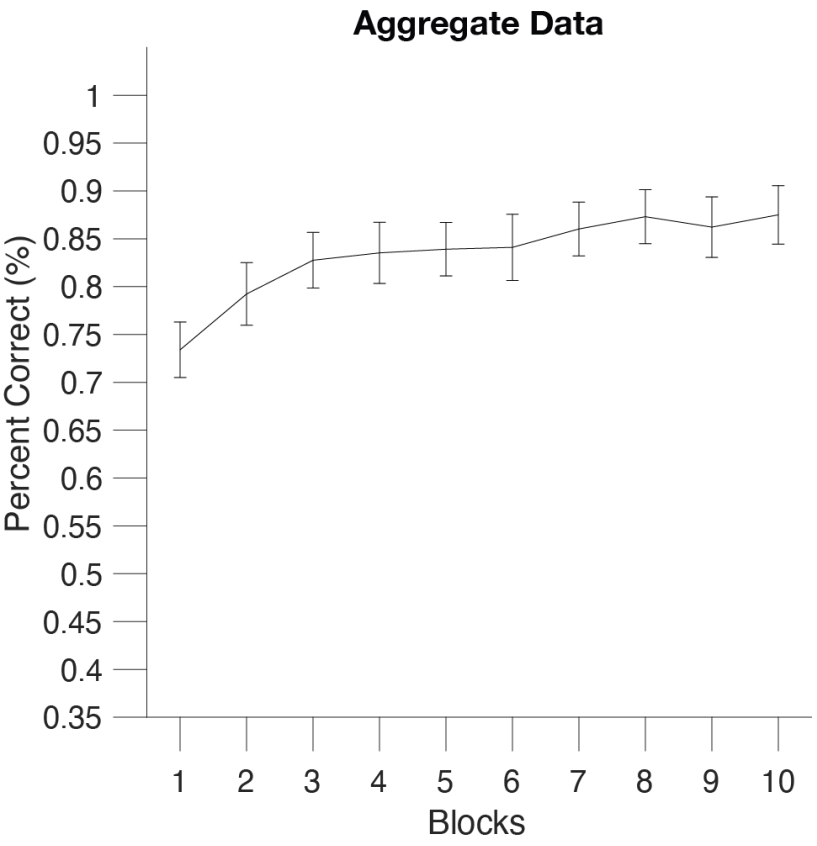


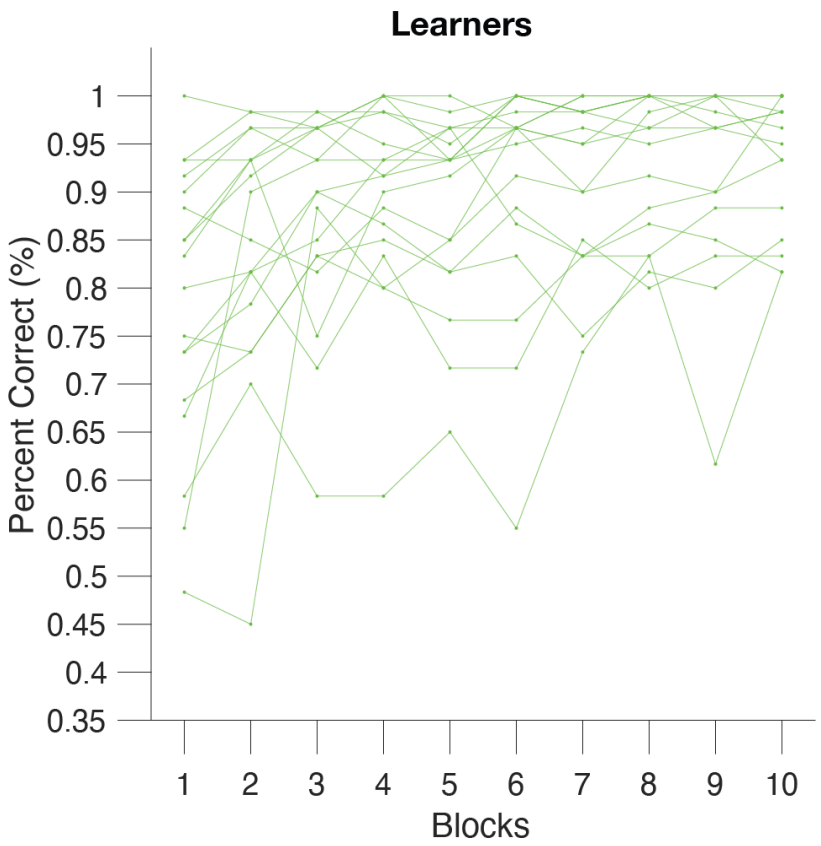
Awareness Test

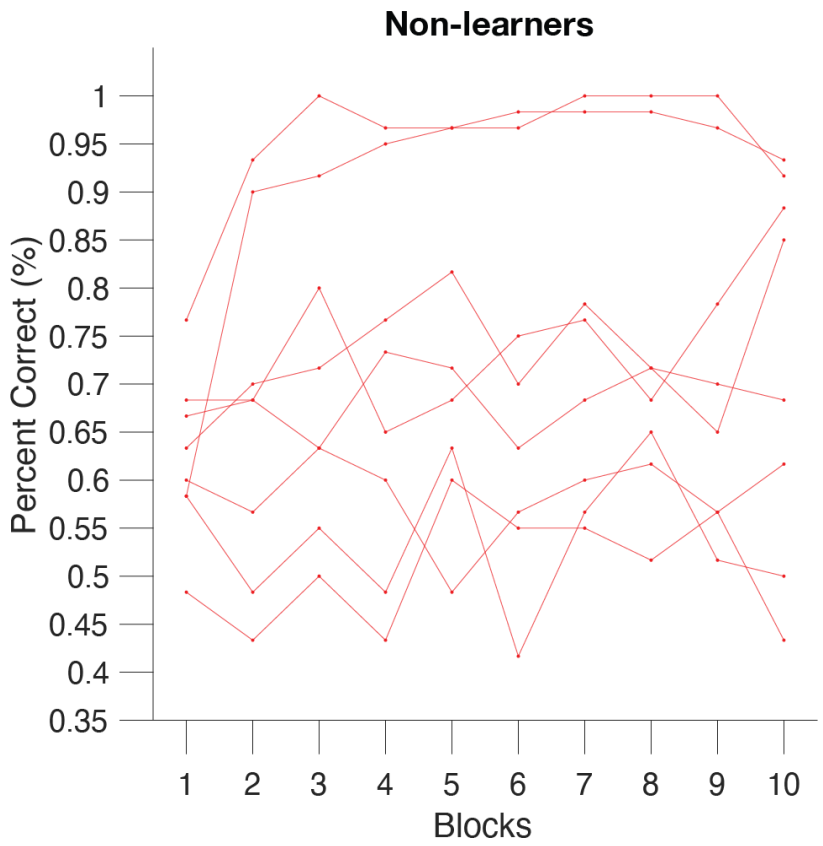
- Only subjects who correctly produced all triplets were considered “learners”



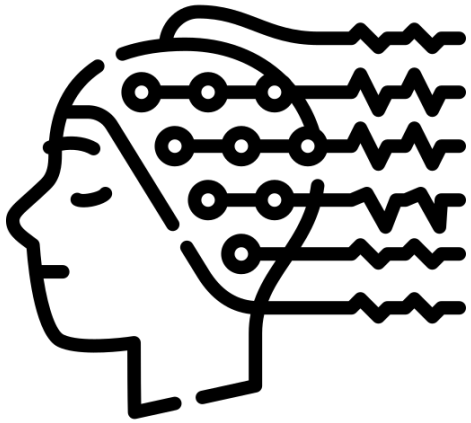
Training Results

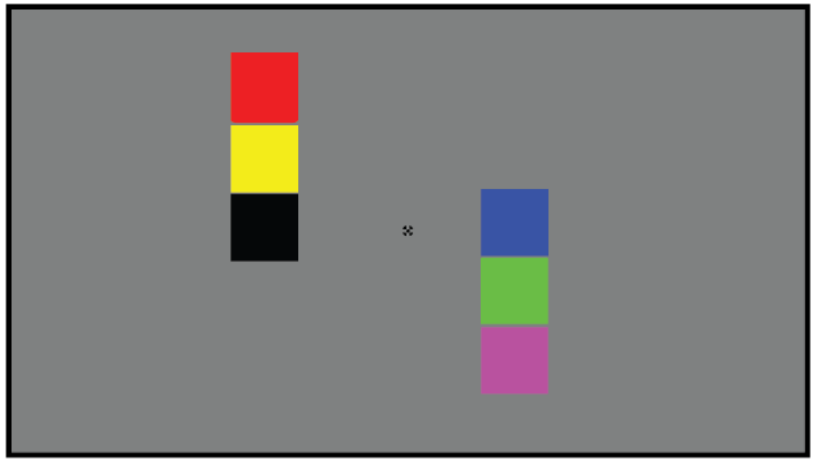


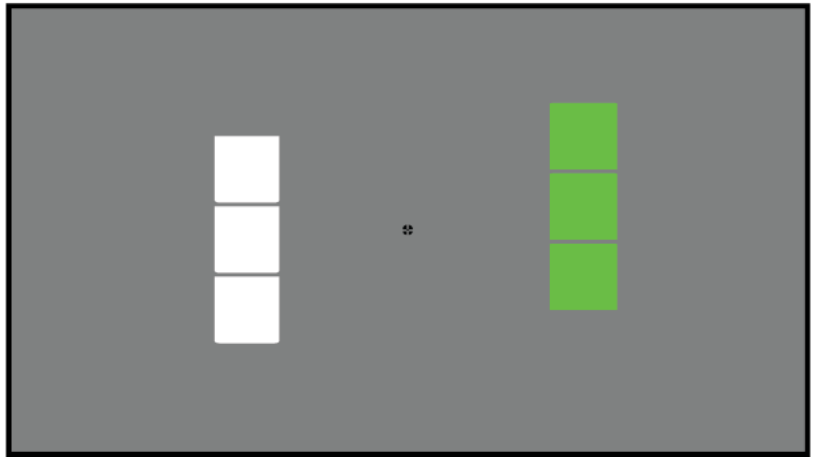




EEG session







Perceptually equivalent

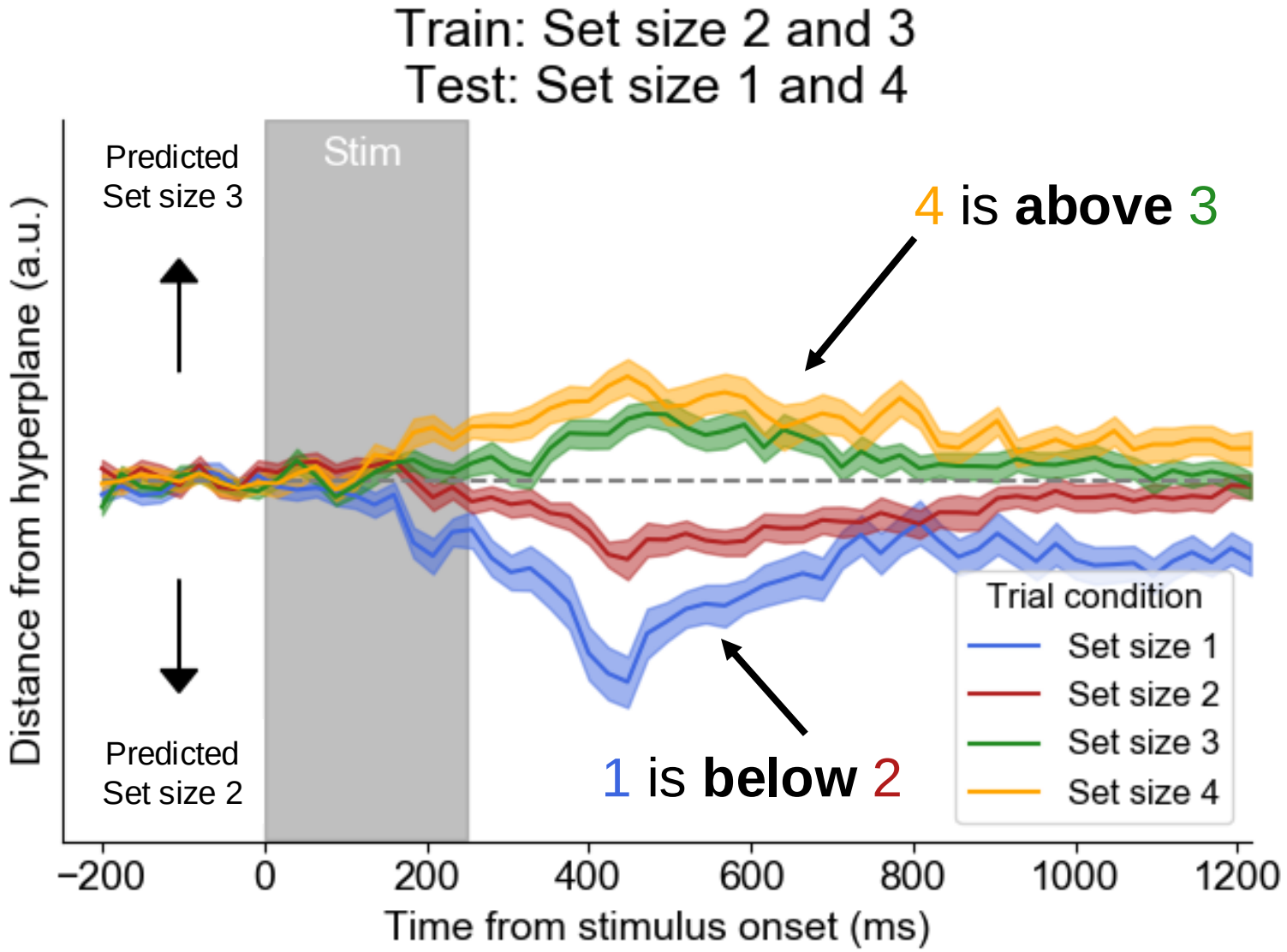
Six random

Six chunked

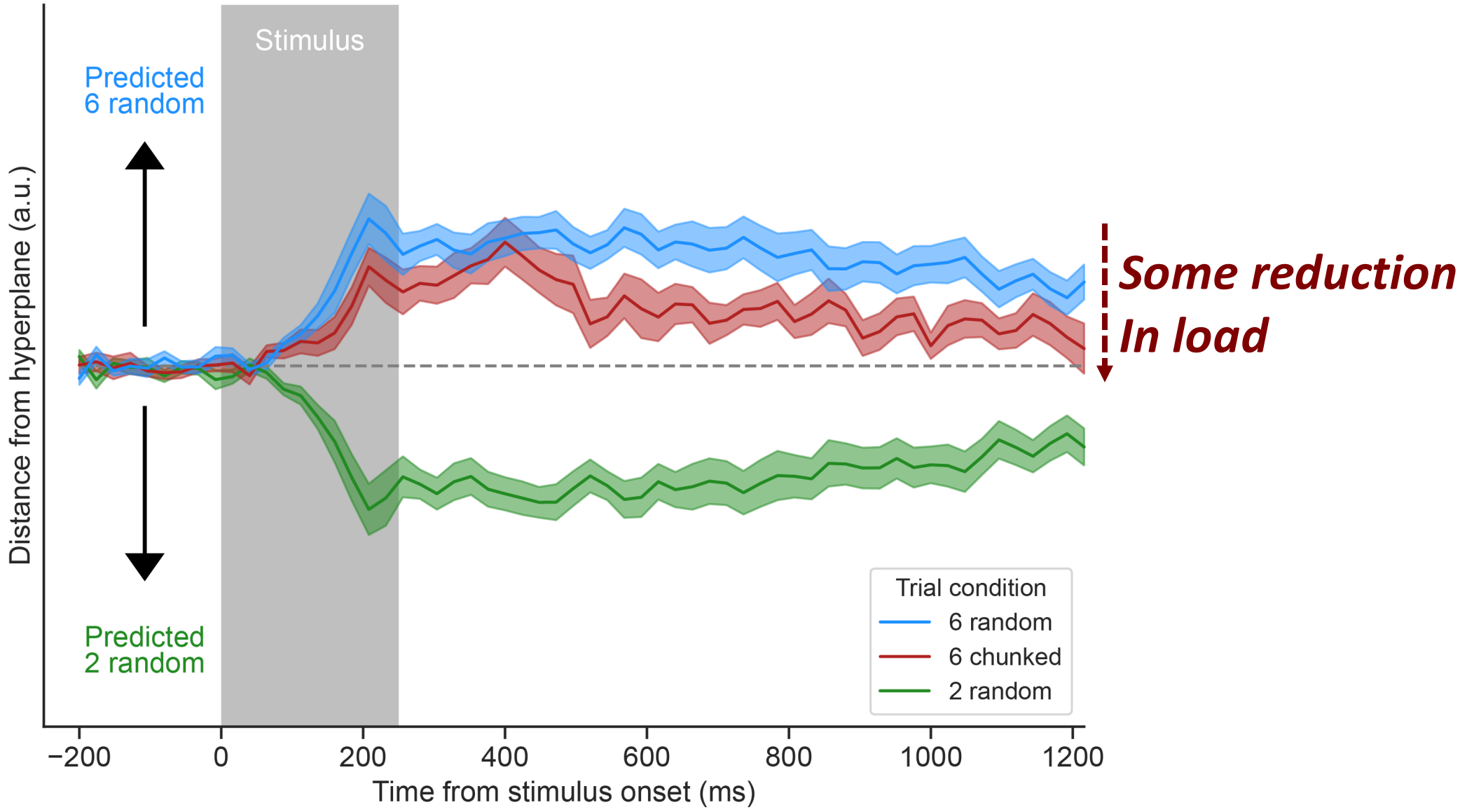
Two random

Multivariate classification of working memory

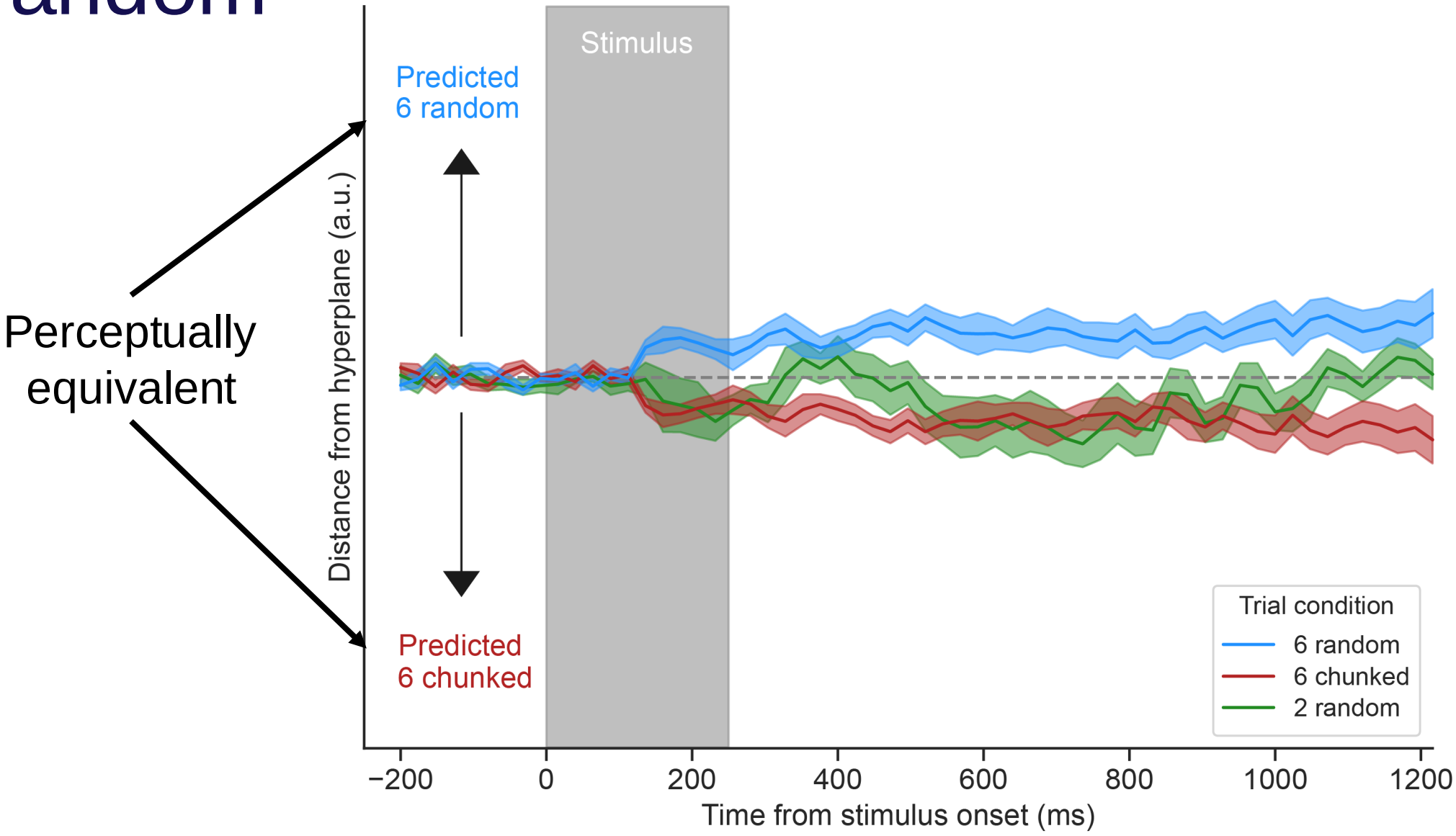
Working memory load signal



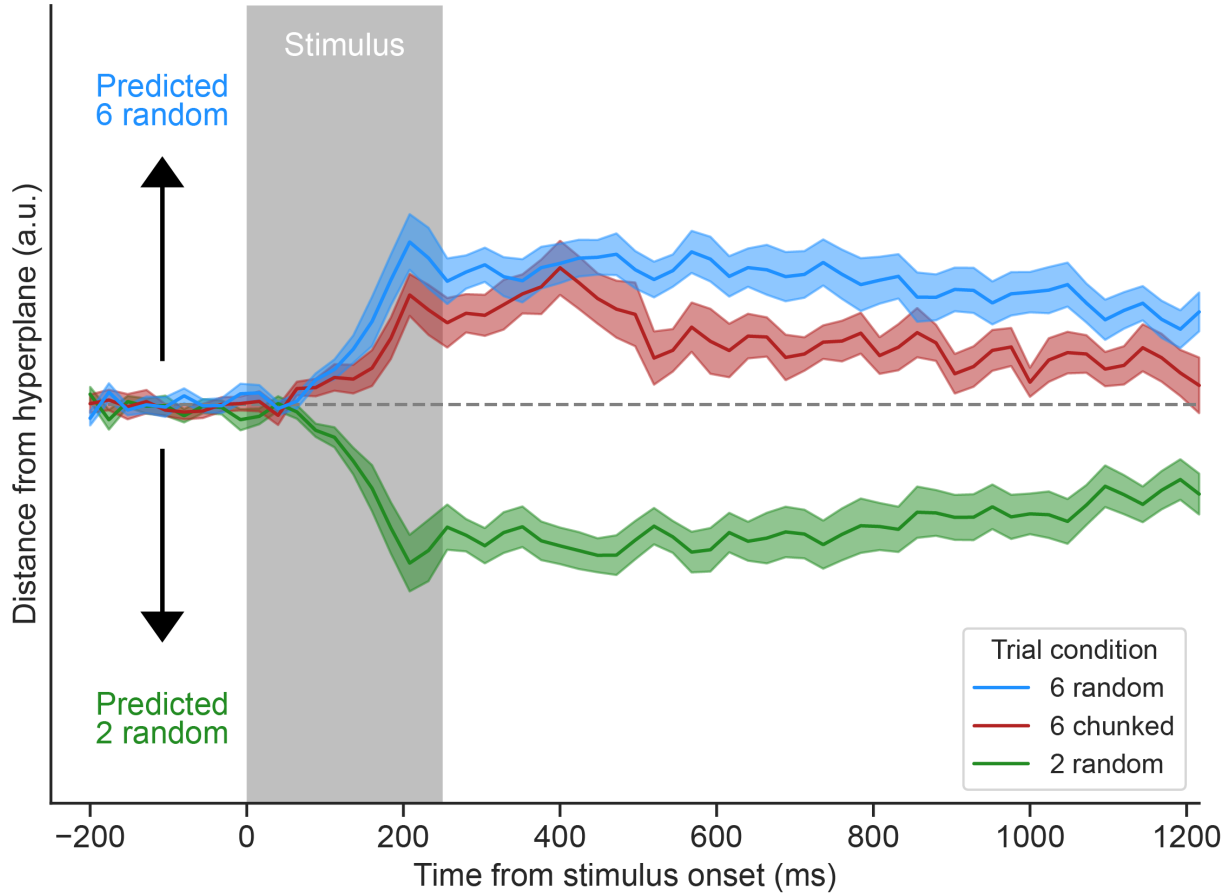
Train 6 random versus 2 random, test 6 chunked



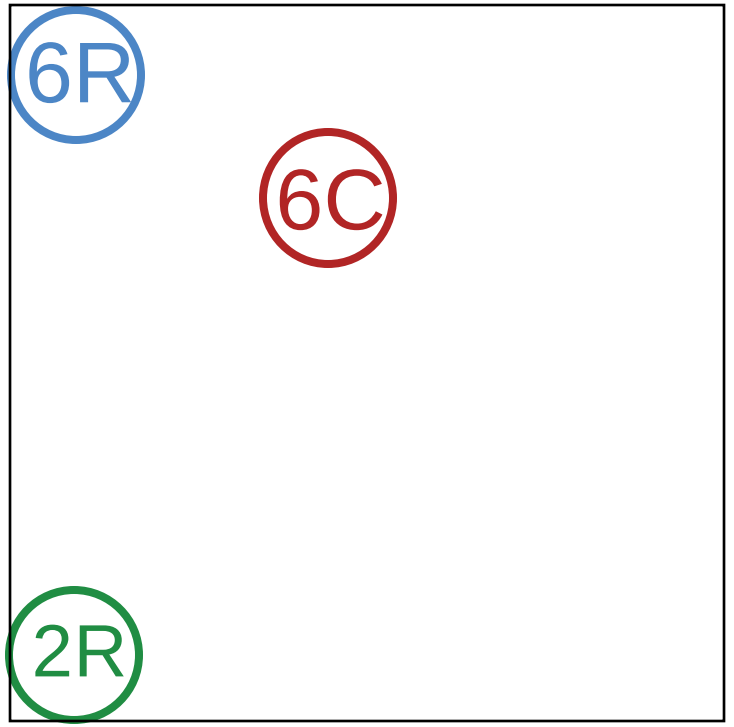
Train 6 random versus 6 chunked, test 2 random



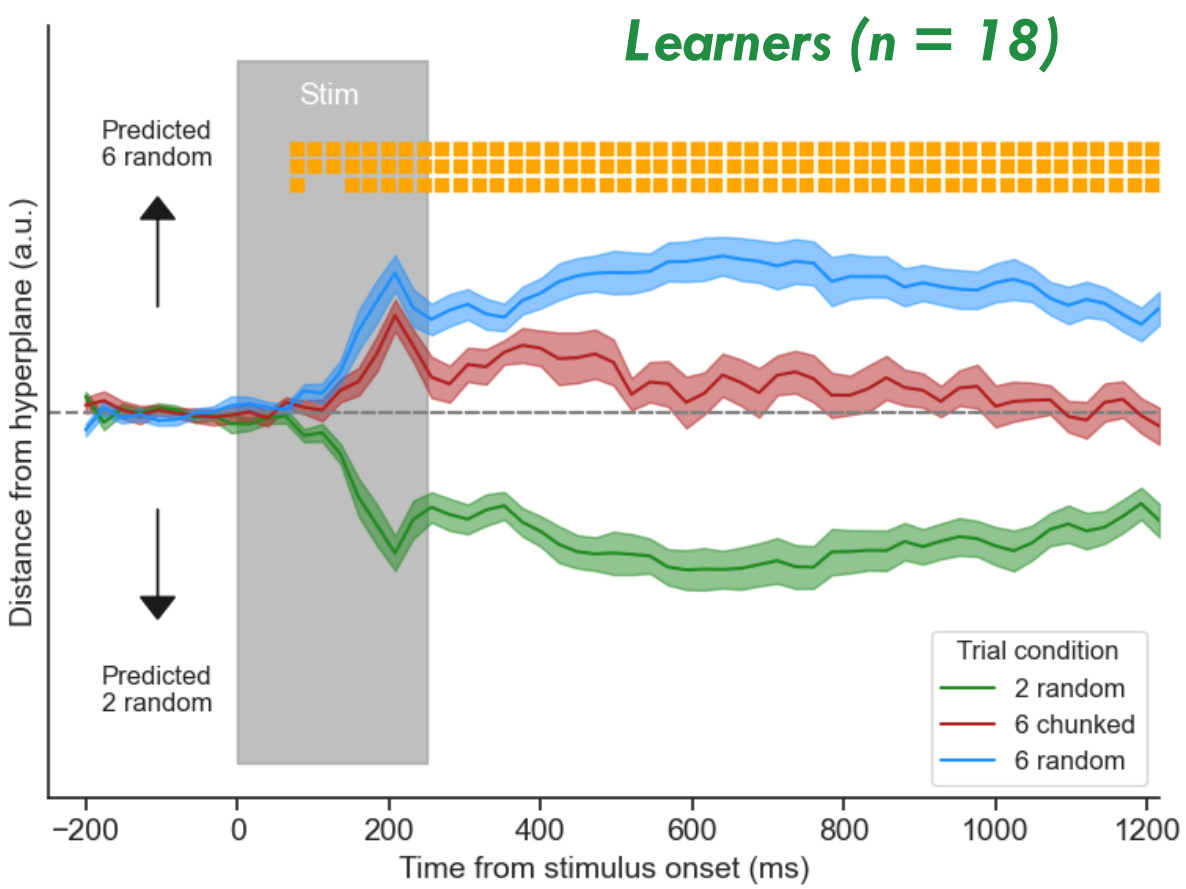
Train 6 random versus 2 random, test 6 chunked

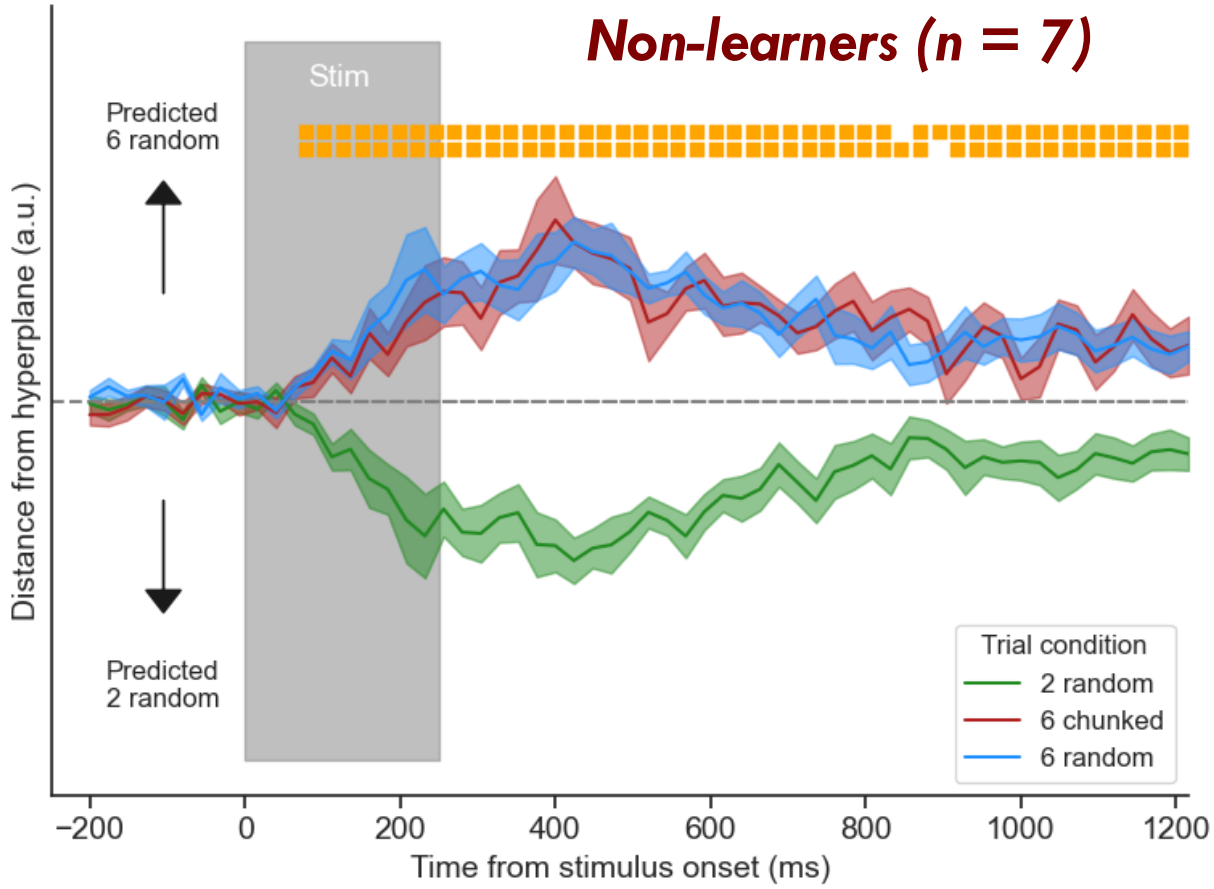


Multidimensional scaling



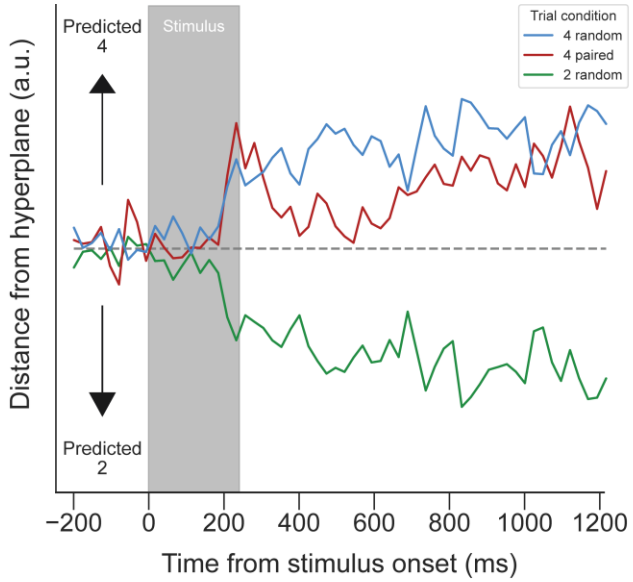
Learners vs non-learners

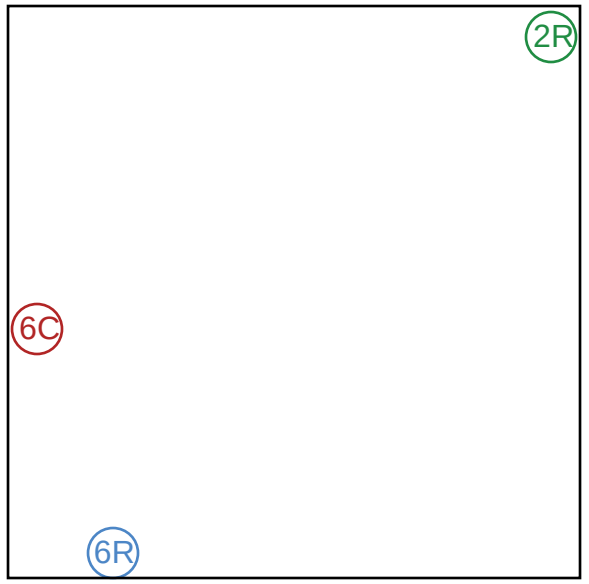




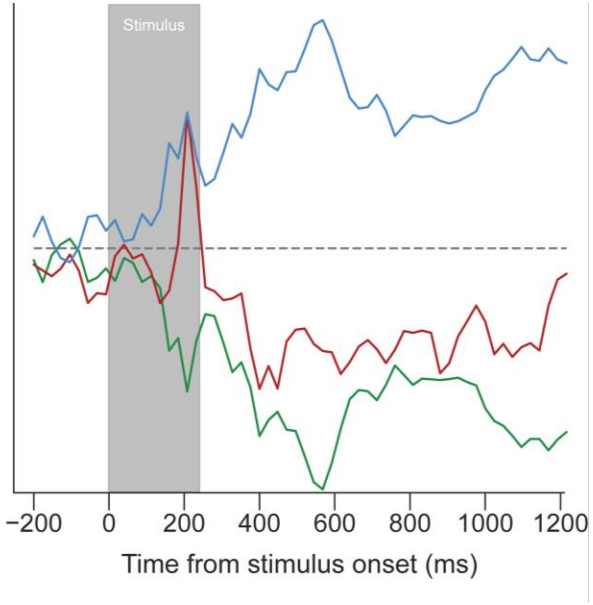
Individual differences

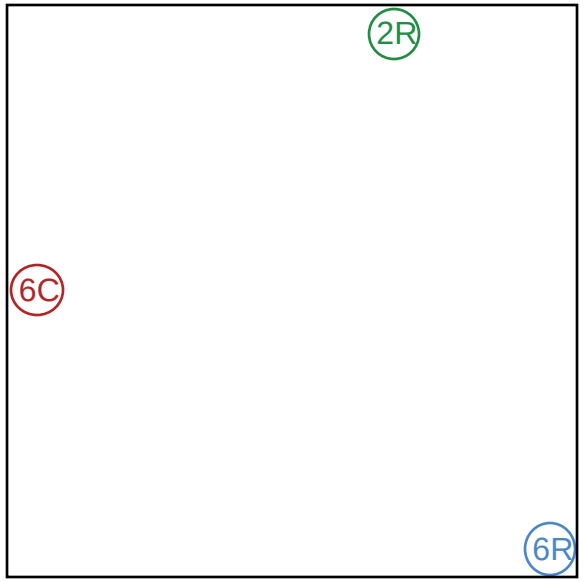
"Weak chunking"





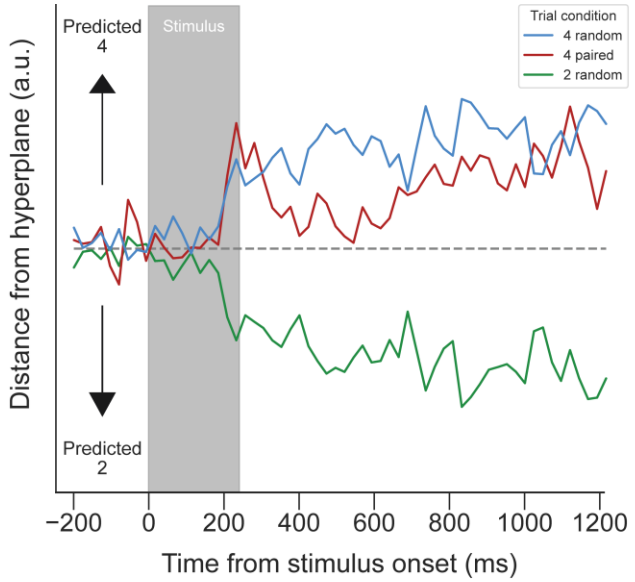
"Strong chunking"

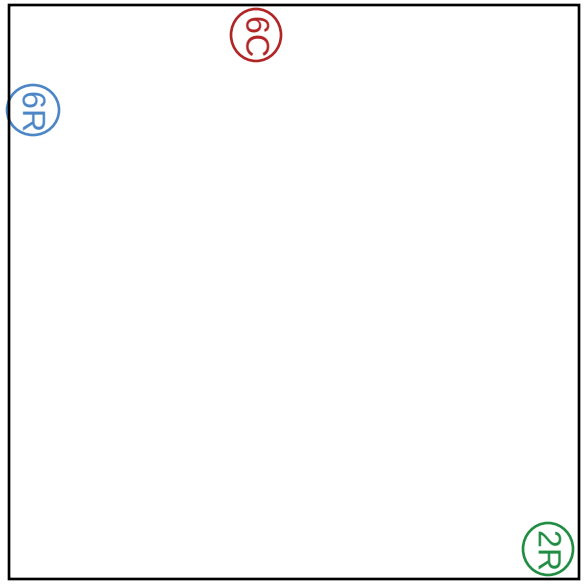




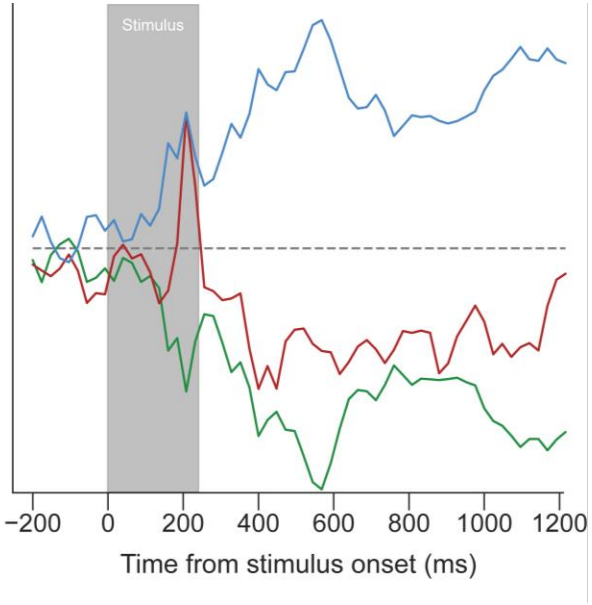
Individual differences

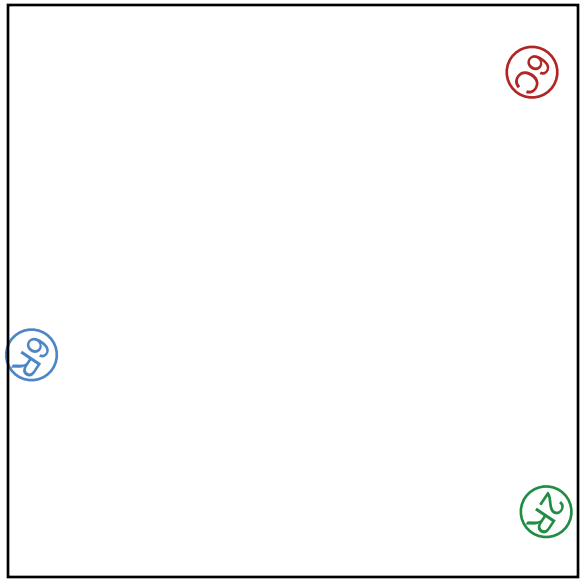
"Weak chunking"





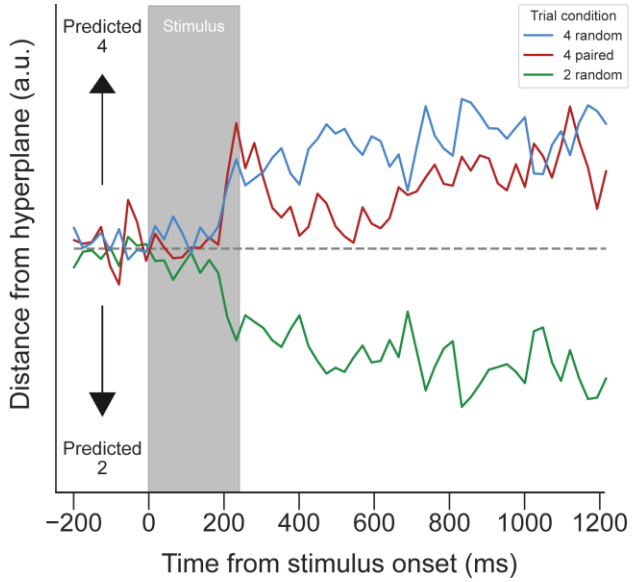
"Strong chunking"

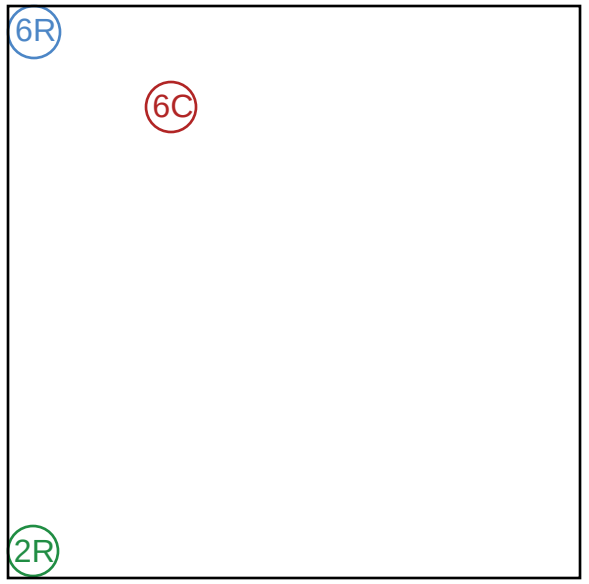




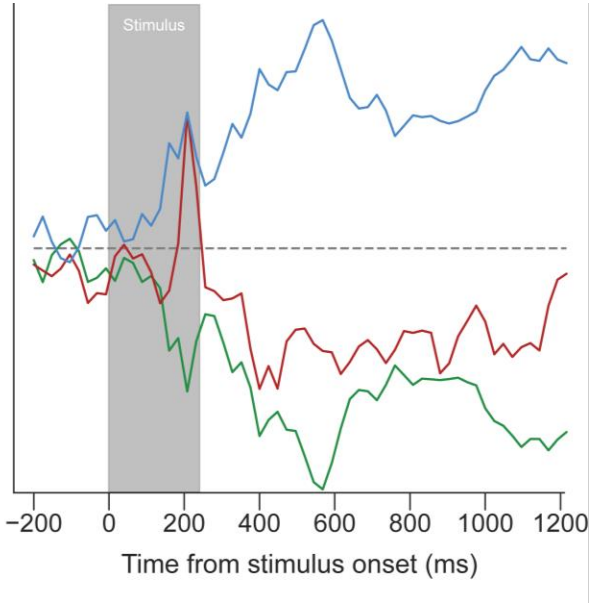
Individual differences

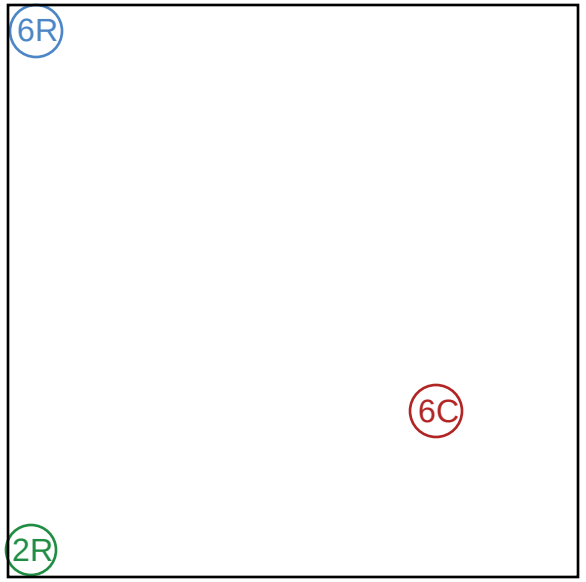
"Weak chunking"





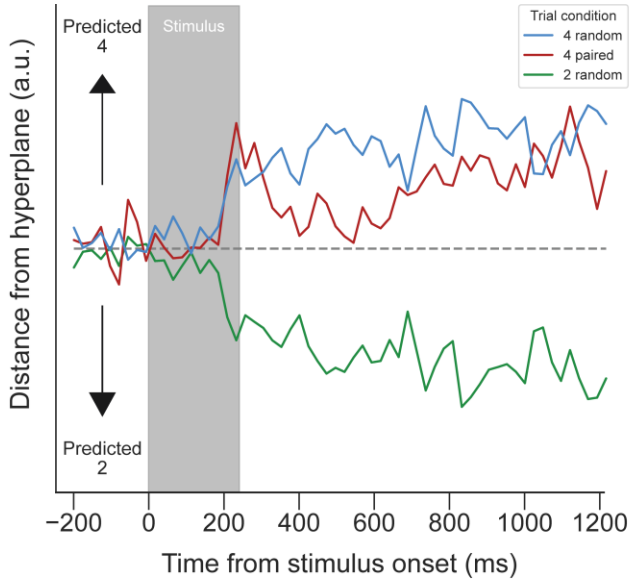
"Strong chunking"



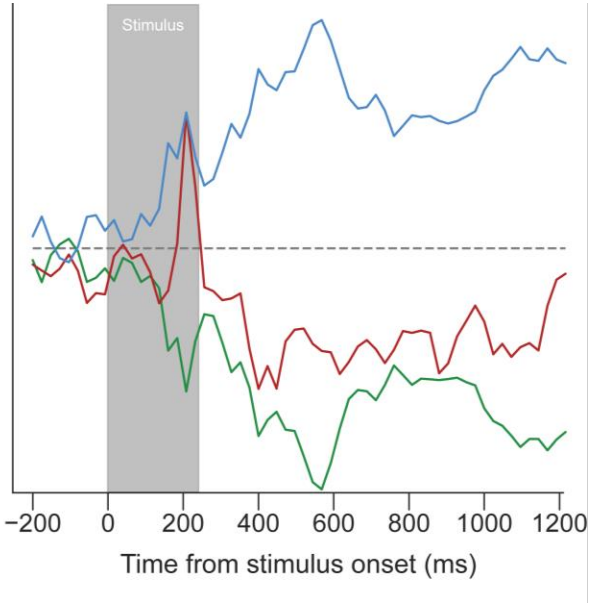


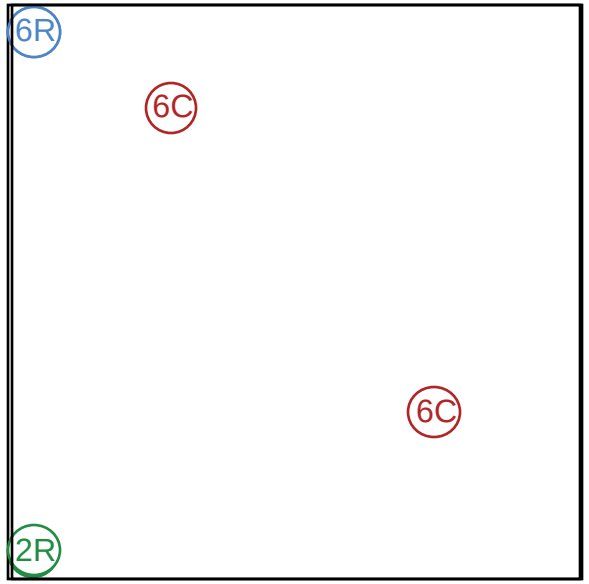
Individual differences

"Weak chunking"

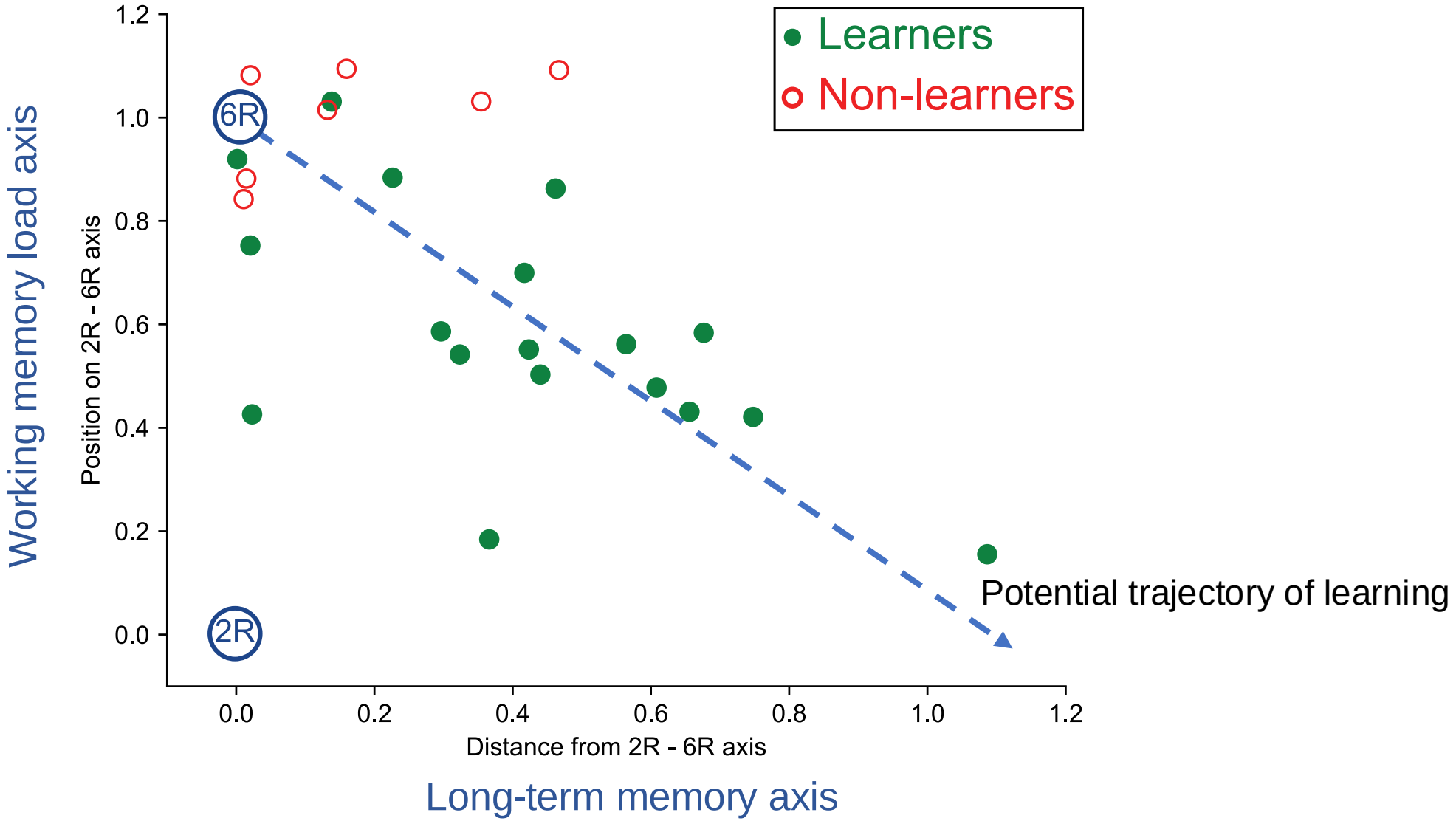


"Strong chunking"





Multidimensional scaling on each subject



Conclusions

- A multivariate neural signal for items in working memory shows associative learning ***reduces*** the number of items stored in working memory

- Furthermore, neural signatures of associative learning showed the reduction only in those that **successfully learnt the associations**

- This is consistent with a ***chunking*** account – associative learning may not allow one to circumvent item pointer limits