Geometric models of the representation underlying working memory

William X. Q. Ngiam

School of Psychology, Adelaide University

In this talk…

1. What is a representation in working memory?
2. Geometric models of working memory performance.
3. Representational geometries in the brain.
4. Aligning representational geometries: worth it or not?

Understanding cognitive limits

I study how information is selected (**attention**) and held in mind (**working memory**) for ongoing perception and cognition.

On the next slide, an image will flicker on and off.

Clap your hands once you have found what is changing in the scene.

Understanding cognitive limits

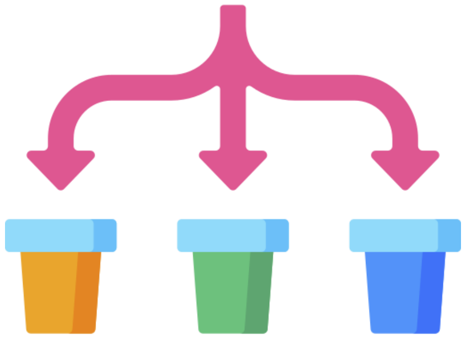
I study how information is selected (**attention**) and held in mind (**working memory**) for ongoing perception and cognition.

One key feature of this system is that it is **capacity-limited.**

Understanding cognitive limits

Object-based theory

"slot models"



versus

Feature-based theory

"resource models"

Luck and Vogel (1997). https://doi.org/10.1038/36846
Zhang and Luck (2008). https://doi.org/10.1038/nature06860

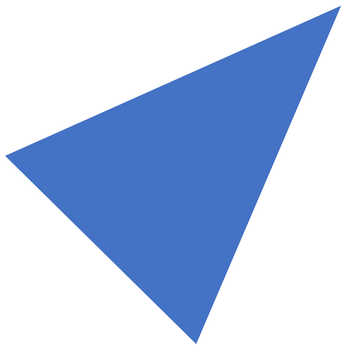
Alvarez and Cavanagh (2004). https://doi.org/10.1111/j.0963-7214.2004.01502006.x
Bays and Husain. (2008) https://doi.org/10.1126/science.1158023

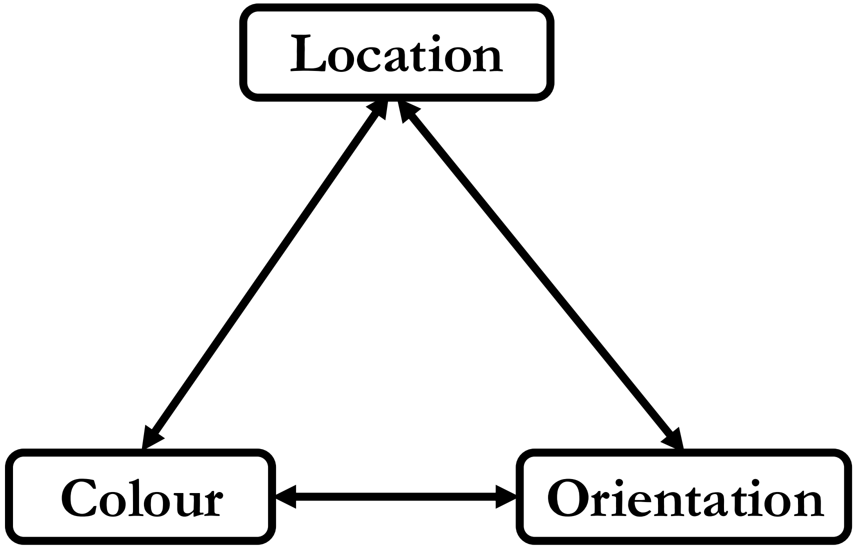
Understanding cognitive limits

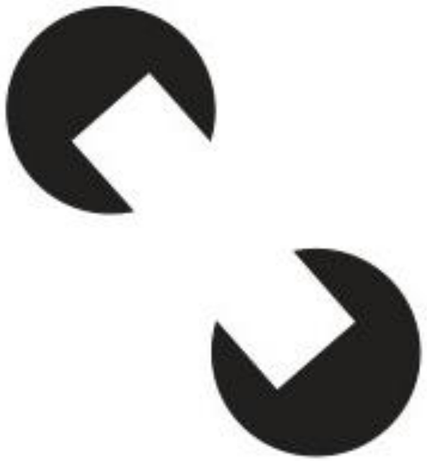
Measuring the capacity limits of visual attention and working memory requires understanding how information is **represented** in mind.

Suchow, J. W., Fougnie, D., Brady, T. F., & Alvarez, G. A. (2014). Terms of the debate on the format and structure of visual memory. *Attention, Perception, & Psychophysics*, *76*(7), 2071-2079. https://doi.org/10.3758/s13414-014-0690-7

How do we represent these in mind and brain?



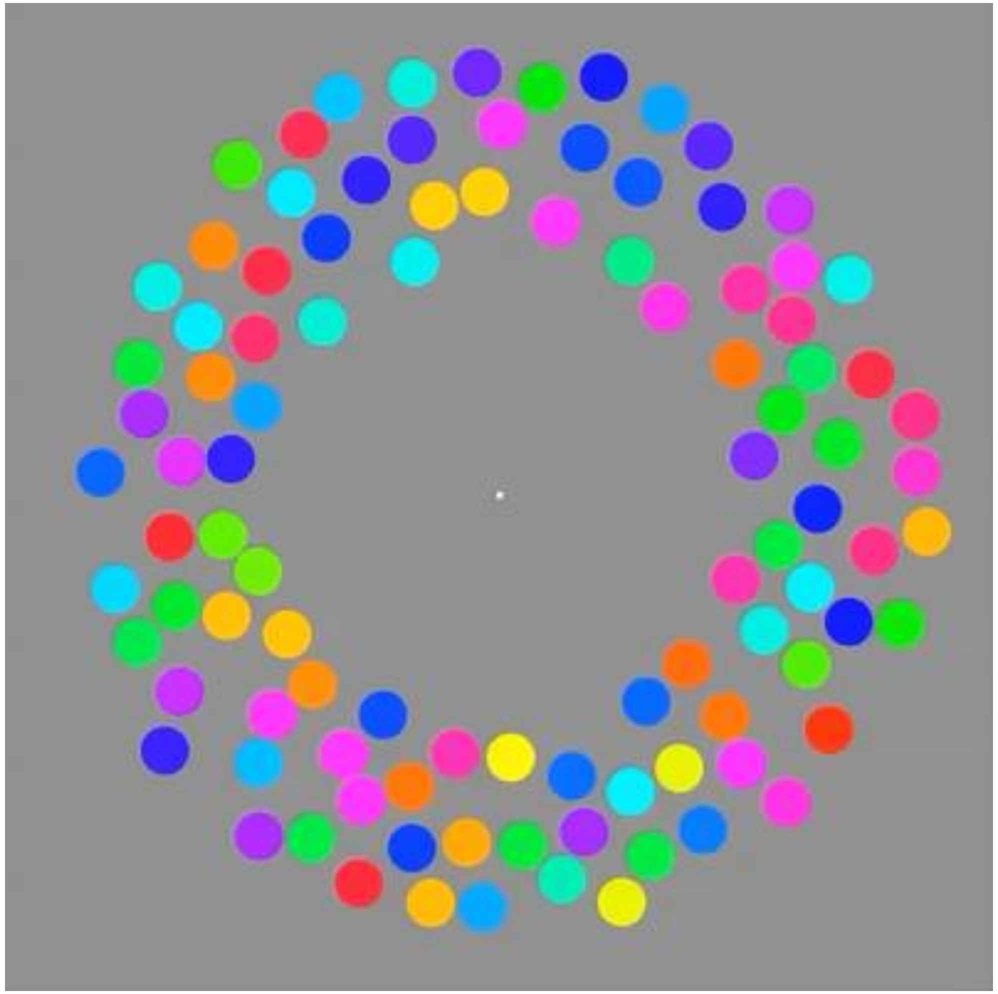


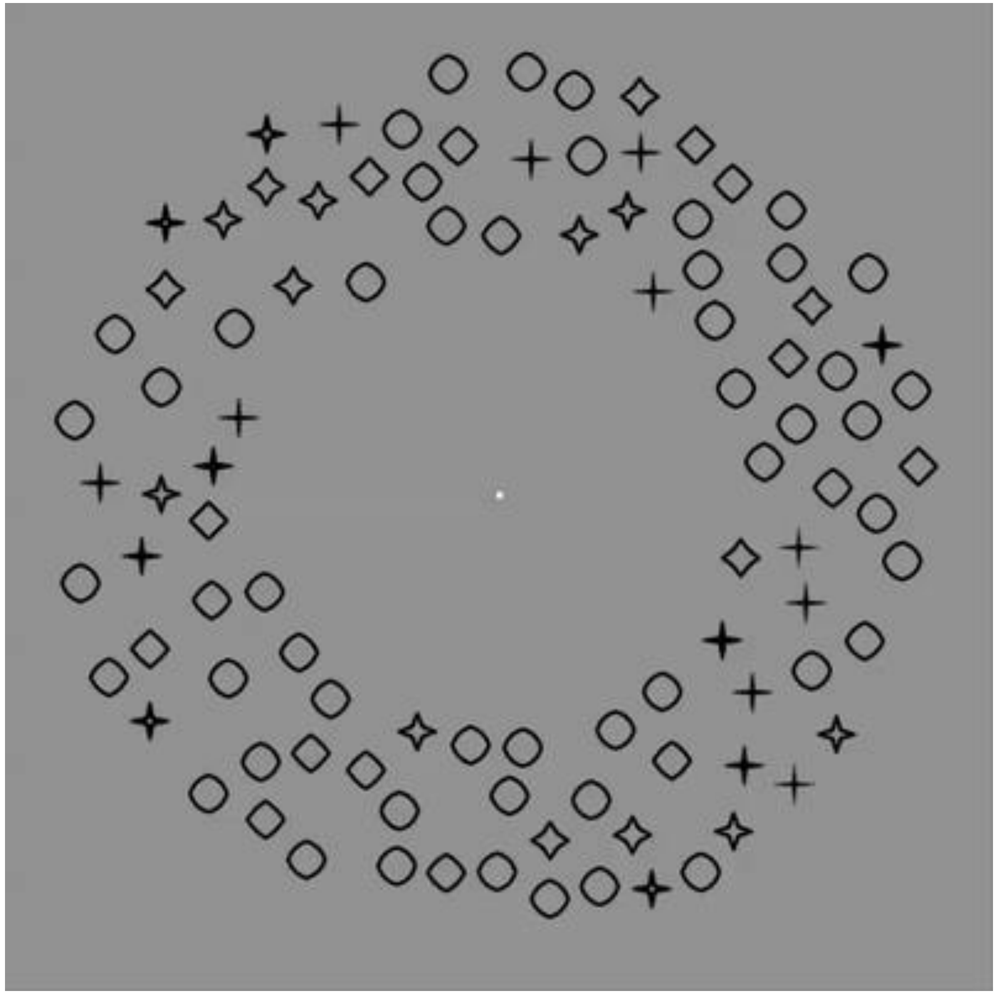


Pieces?
Illusory
objects?

Memories across
space and time?

What is being represented in mind?





Video from Jordan Suchow's YouTube channel: https://www.youtube.com/watch?v=-KhPYdge9RU

Suchow, J. W., & Alvarez, G. A. (2011). Motion silences awareness of visual change. *Current Biology*, *21*(2), 140-143.

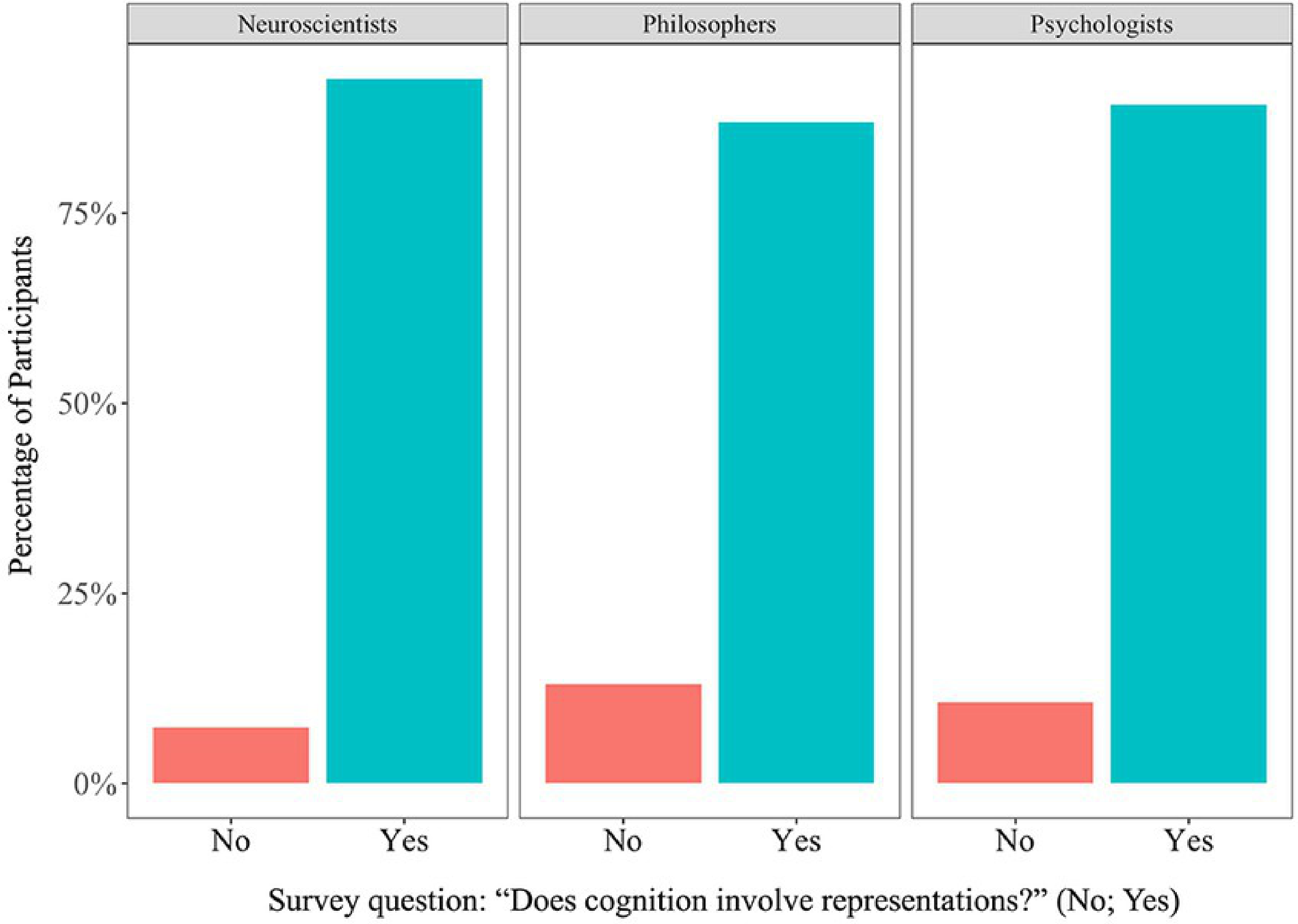
Understanding cognitive limits

Measuring the capacity limits of visual attention and working memory requires understanding how information is **represented** in mind.

But **what is a representation?**

What is a representation?

Favela and Machery (2023) surveyed 736 neuroscientists, philosophers and psychologists about representations.



Favela, L. H., & Machery, E. (2023). Investigating the concept of representation in the neural and psychological sciences. Frontiers in Psychology, 14, 1165622. 10.3389/fpsyg.2023.1165622

Favela, L. H., & Machery, E. (2025). The concept of representation in the brain sciences: The current status and ways forward. *Mind & Language*, *40*(2), 215-225. 10.1111/mila.12531

What is a representation?

They find that the concept of representation is **applied inconsistently** by psychologists, neuroscientists and philosophers.

- Disagreement on whether the brain's response counts as representing the stimulus
- Imprecise about what brain structure or pattern counts as representation (i.e. single neuron or population)
- Reluctance for failure of the brain to count as misrepresenting

Favela, L. H., & Machery, E. (2023). Investigating the concept of representation in the neural and psychological sciences. Frontiers in Psychology, 14, 1165622. 10.3389/fpsyg.2023.1165622
Favela, L. H., & Machery, E. (2025). The concept of representation in the brain sciences: The current status and ways forward. *Mind & Language*, *40*(2), 215-225. 10.1111/mila.12531

What is a representation?

Favela and Machery (2025) say that the discordance about representations is useful; that the concept does not need reform or to be eliminated to be more clear.

Rosa Cao (2025) argues for the **pragmatic view** – that it only matters if we can identify the "representation" and manipulate

Frances Egan (2025) argues that neuroscientists or psychologists actually mean "**causes or carries information about**" when relying on representation terms.

Favela, L. H., & Machery, E. (2023). Investigating the concept of representation in the neural and psychological sciences. Frontiers in Psychology, 14, 1165622. 10.3389/fpsyg.2023.1165622
Favela, L. H., & Machery, E. (2025). The concept of representation in the brain sciences: The current status and ways forward. *Mind & Language*, *40*(2), 215-225. 10.1111/mila.12531
Cao, R. (2022). Putting representations to use. Synthese, 200(2), 151. 10.1007/s11229-022-03522-3
Egan, F. (2025). Comments on Favela and Machery's The concept of representation in the brain sciences: The current status and ways forward. Mind & Language, 40(2), 233-238. 10.1111/mila.12527

What is a representation?

It is clear that there is no single definition of representation that we can point to; and perhaps representation should be the **explanans** (the thing to be explained) rather than the explanandum.

I will present two recent projects on **representational geometries** using different approaches:

1. Mathematical modeling of cognitive behaviour

2. Machine learning classification of neuroimaging data

The representation in classic cognitive models

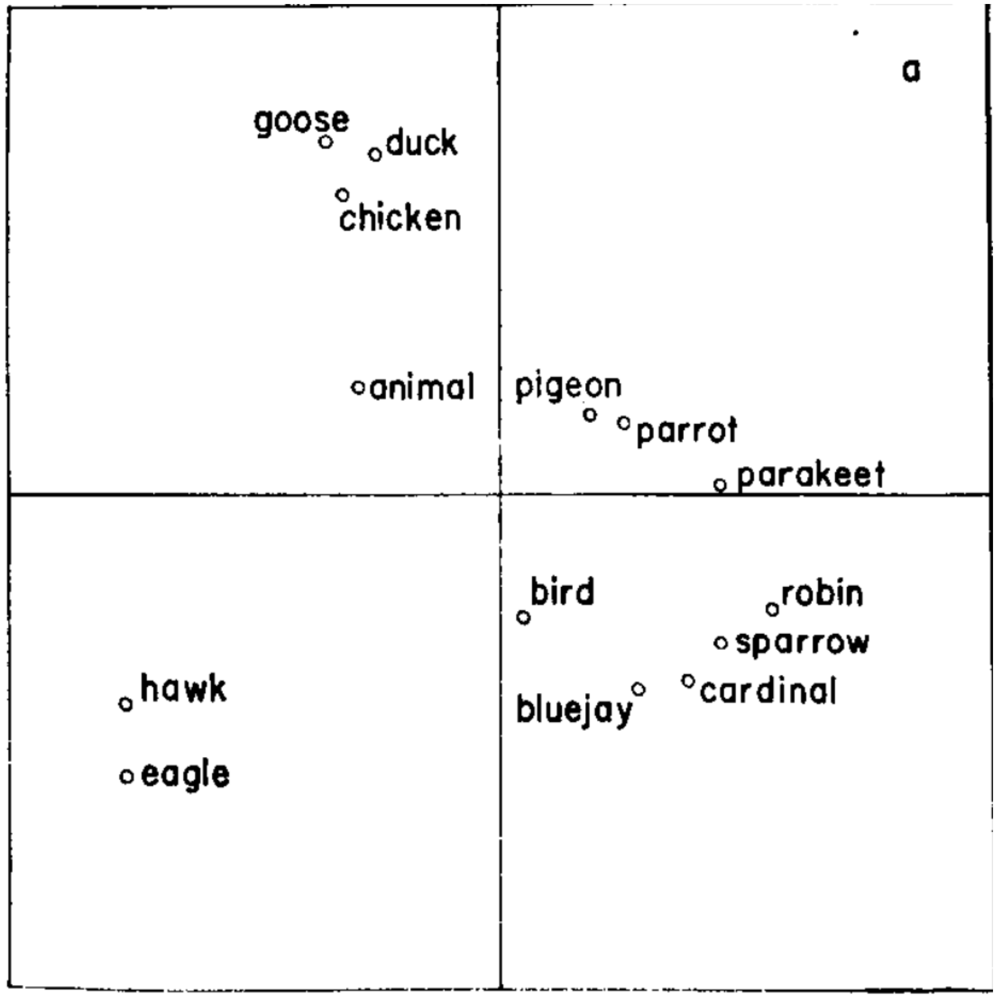
Classic formal models of cognition would derive the **psychological representation** using multidimensional scaling (MDS) of similarity judgments.

- Collect similarity ratings for pairs of items in the stimulus set
- Reduce those ratings into a representation that best preserves the distance between the items; the closer the items, the more similar.

The representation in classic cognitive models

This is the MDS representation for **birds**.

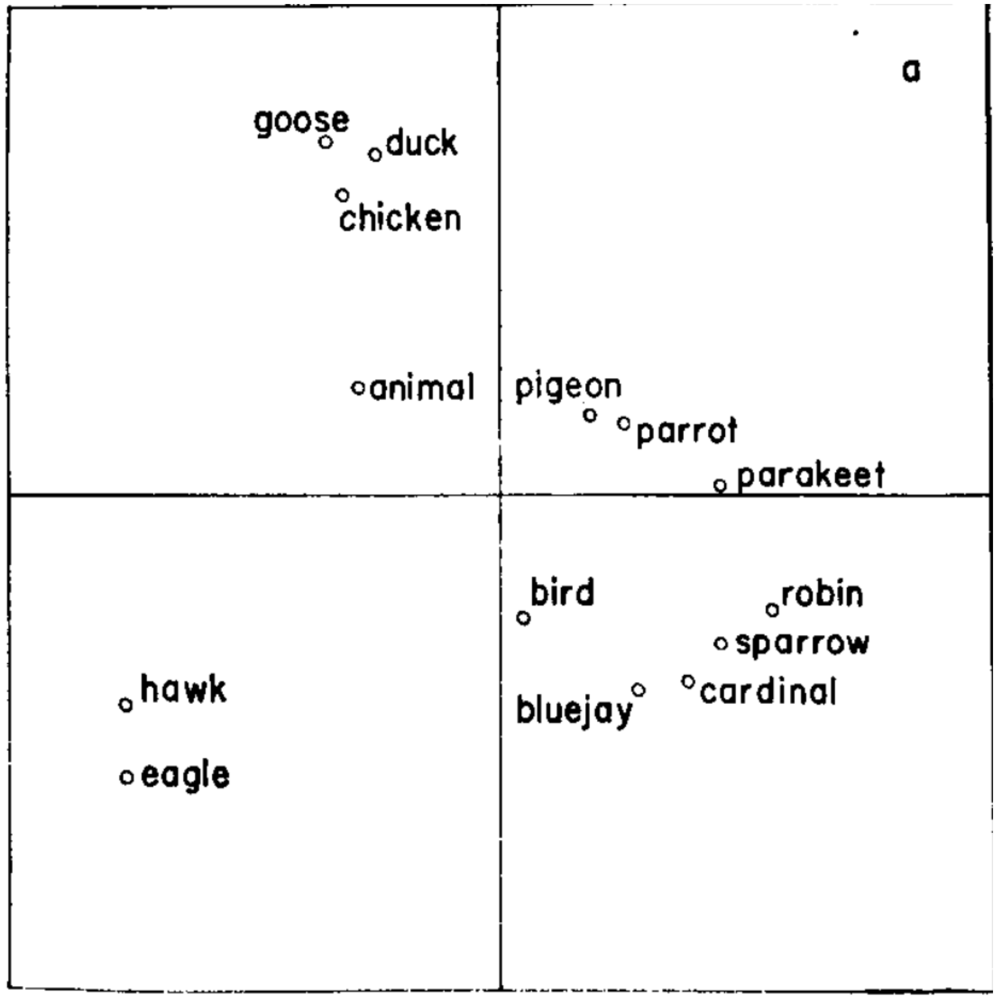
The distances between the birds reflect their psychological similarity; the closer, the more similar.



Rips, L. J., Shoben, E. J., & Smith, E. E. (1973). Semantic distance and the verification of semantic relations. *Journal of verbal learning and verbal behavior*, *12*(1), 1-20.

The representation in classic cognitive models

Formal models use these distances to predict behaviour (such as induction and generalisation).



Rips, L. J., Shoben, E. J., & Smith, E. E. (1973). Semantic distance and the verification of semantic relations. *Journal of verbal learning and verbal behavior*, *12*(1), 1-20.

The representation in classic cognitive models

Formal models use these distances to predict behaviour (such as induction and generalisation).

For example, say **ducks** are susceptible to Reinholf disease…

Rips, L. J., Shoben, E. J., & Smith, E. E. (1973). Semantic distance and the verification of semantic relations. *Journal of verbal learning and verbal behavior*, *12*(1), 1-20.

The representation in classic cognitive models

Formal models use these distances to predict behaviour (such as induction and generalisation).

For example, say **ducks** are susceptible to Reinholf disease… do you think **chickens** are also susceptible?

Rips, L. J., Shoben, E. J., & Smith, E. E. (1973). Semantic distance and the verification of semantic relations. *Journal of verbal learning and verbal behavior*, *12*(1), 1-20.

The representation in classic cognitive models

Formal models use these distances to predict behaviour (such as induction and generalisation).

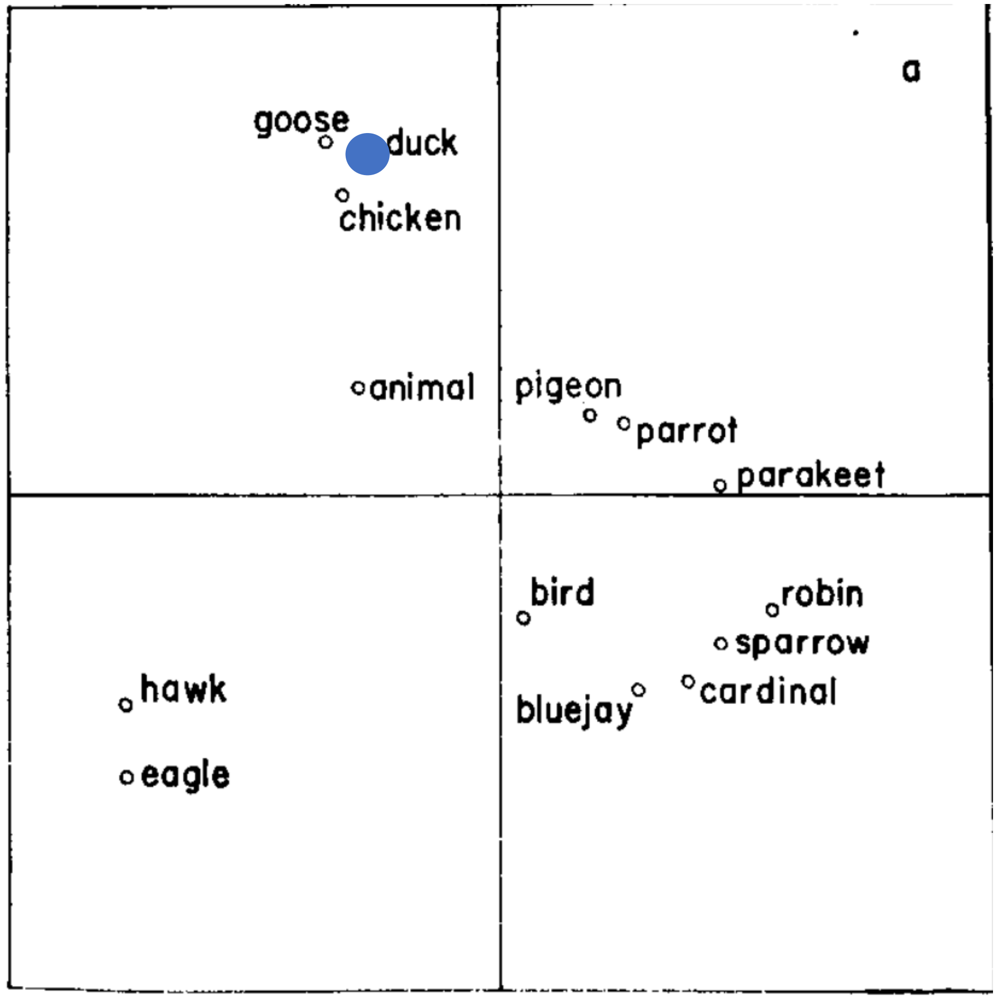
For example, say **ducks** are susceptible to Reinholf disease… do you think **hawks** are also susceptible?

Rips, L. J., Shoben, E. J., & Smith, E. E. (1973). Semantic distance and the verification of semantic relations. *Journal of verbal learning and verbal behavior*, *12*(1), 1-20.

The representation in classic cognitive models

Formal models use these distances to predict behaviour (such as induction and generalisation).

For example, say **ducks** are susceptible to Reinholf disease…

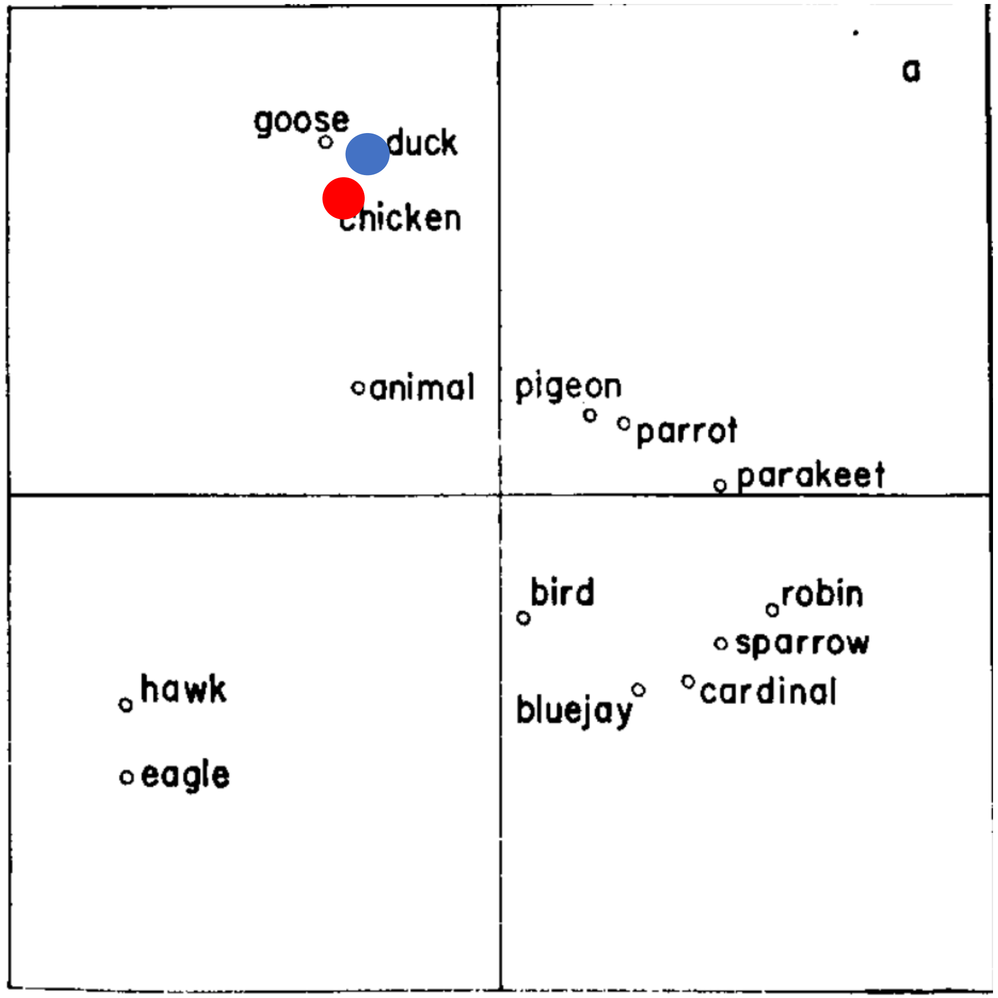


Rips, L. J., Shoben, E. J., & Smith, E. E. (1973). Semantic distance and the verification of semantic relations. *Journal of verbal learning and verbal behavior*, *12*(1), 1-20.

The representation in classic cognitive models

Formal models use these distances to predict behaviour (such as induction and generalisation).

For example, say **ducks** are susceptible to Reinholf disease… one is likely to think **chickens** are also susceptible.

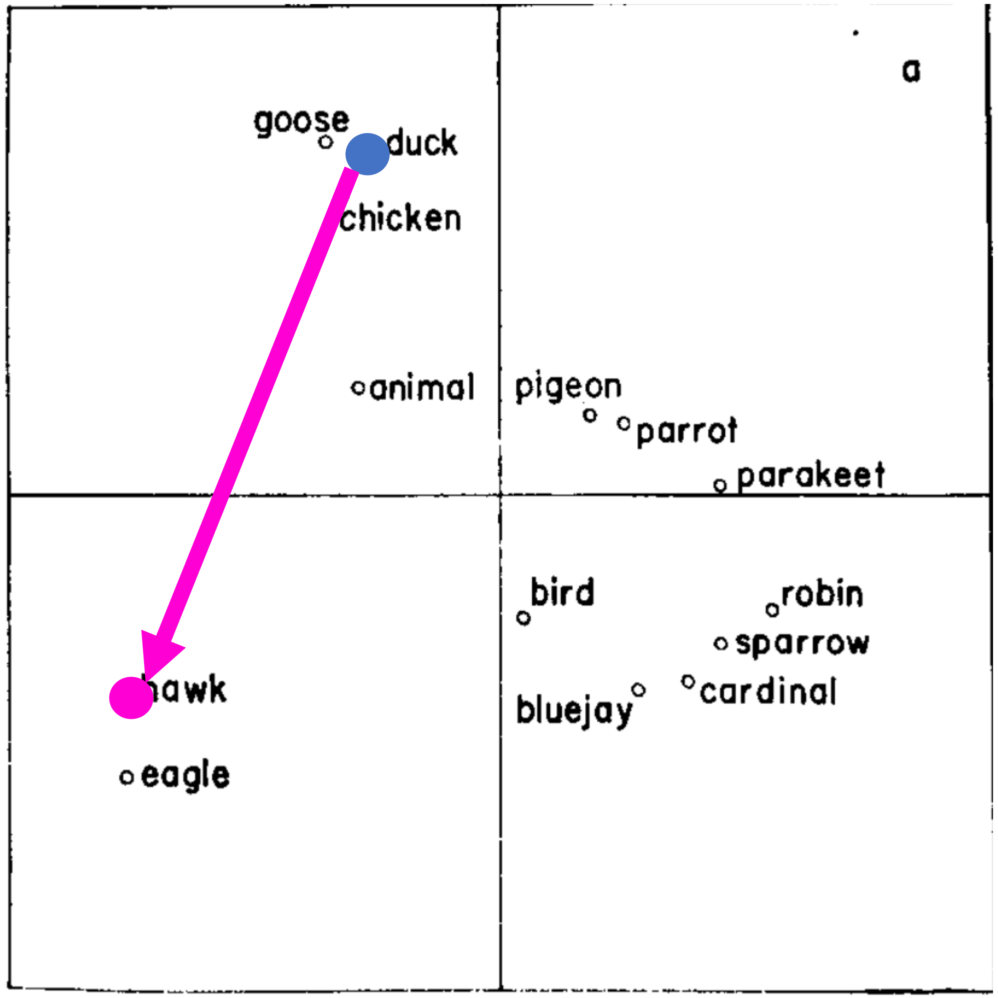


Rips, L. J., Shoben, E. J., & Smith, E. E. (1973). Semantic distance and the verification of semantic relations. *Journal of verbal learning and verbal behavior*, *12*(1), 1-20.

The representation in classic cognitive models

Formal models use these distances to predict behaviour (such as induction and generalisation).

For example, say **ducks** are susceptible to Reinholf disease… but less likely to think **hawks** are also susceptible.



Rips, L. J., Shoben, E. J., & Smith, E. E. (1973). Semantic distance and the verification of semantic relations. *Journal of verbal learning and verbal behavior*, *12*(1), 1-20.

The representation in classic cognitive models

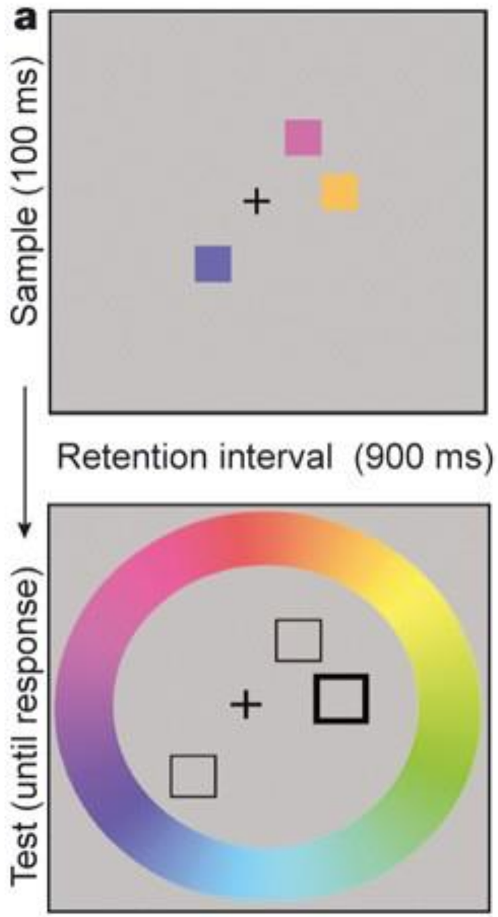
For historical reasons, the similarity-based MDS-representation has been considered **the psychological representation** underlying cognition.

Many cognitive modelers focus on the operations that occur on the representation, rather than explain the representation itself.

Nosofsky, R. M. (1992). Similarity scaling and cognitive process models. *Annual review of Psychology*, *43*(1), 25-53.
Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science*, *237*(4820), 1317-1323.

Two classes of visual working memory models

Most visual working memory models assume **the physical stimulus space** as the representation that cognitive mechanisms operate upon.

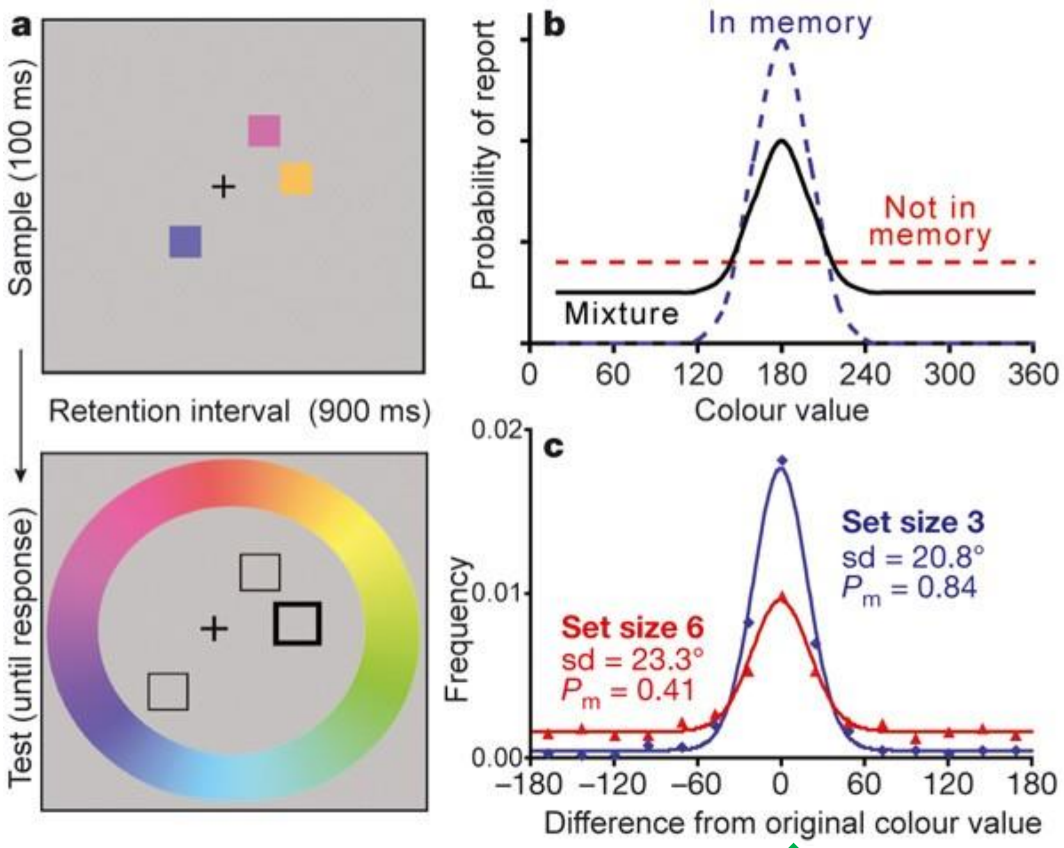


Models assume the physical space, such as this colour wheel

Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature*, *453*(7192), 233-235.

Two classes of visual working memory models

Most visual working memory models assume **the physical stimulus space** as the representation that cognitive mechanisms operate upon.

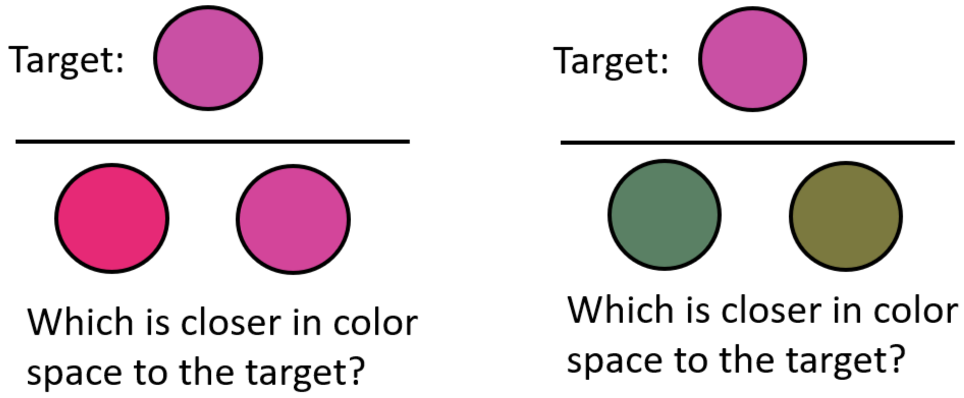


Memory operationalised in terms of absolute error on colour wheel

Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature*, *453*(7192), 233-235.

The representation in visual working memory

One model that (sort of) accounts for conceptual representations is the **Target Confusability Competition (TCC)** model.

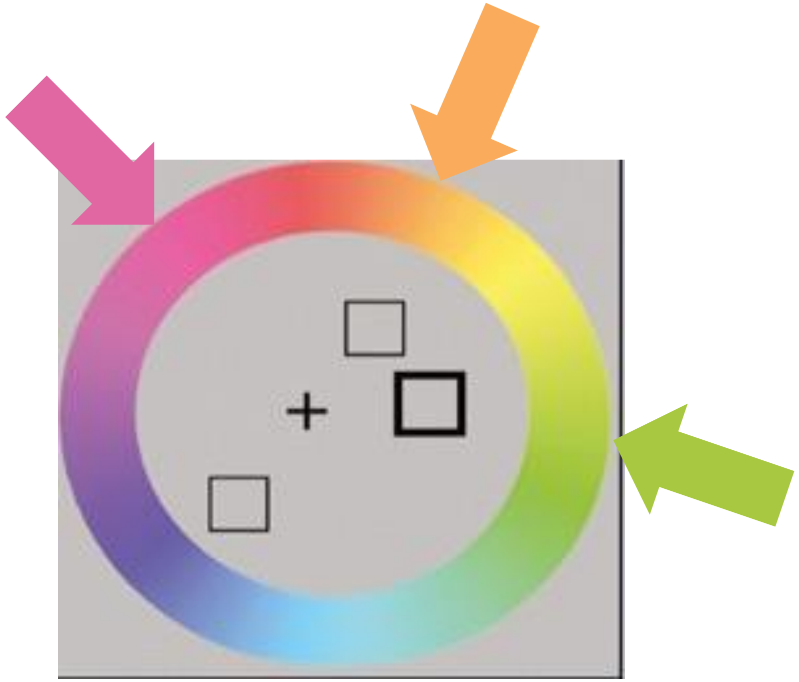


Ekman (1954). https://doi.org/10.1080/00223980.1954.9712953
Schurgin, Wixted and Brady (2020) https://doi.org/10.1038/s41562-020-00938-0

The representation in visual working memory

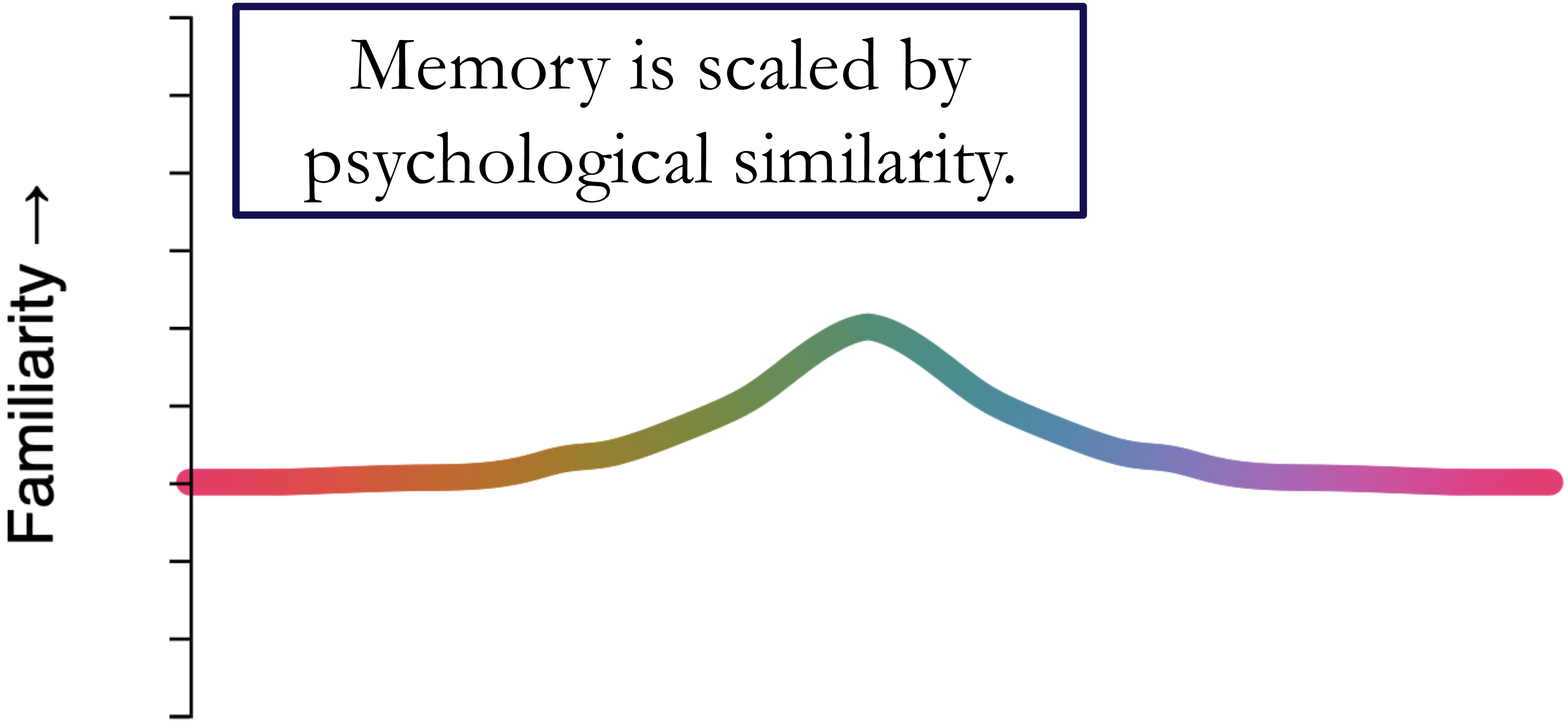
They made the point that confusability would not be captured on the physical stimulus space.

For example, orange is more similar to pink than green, despite being the same distance in absolute degrees.



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.

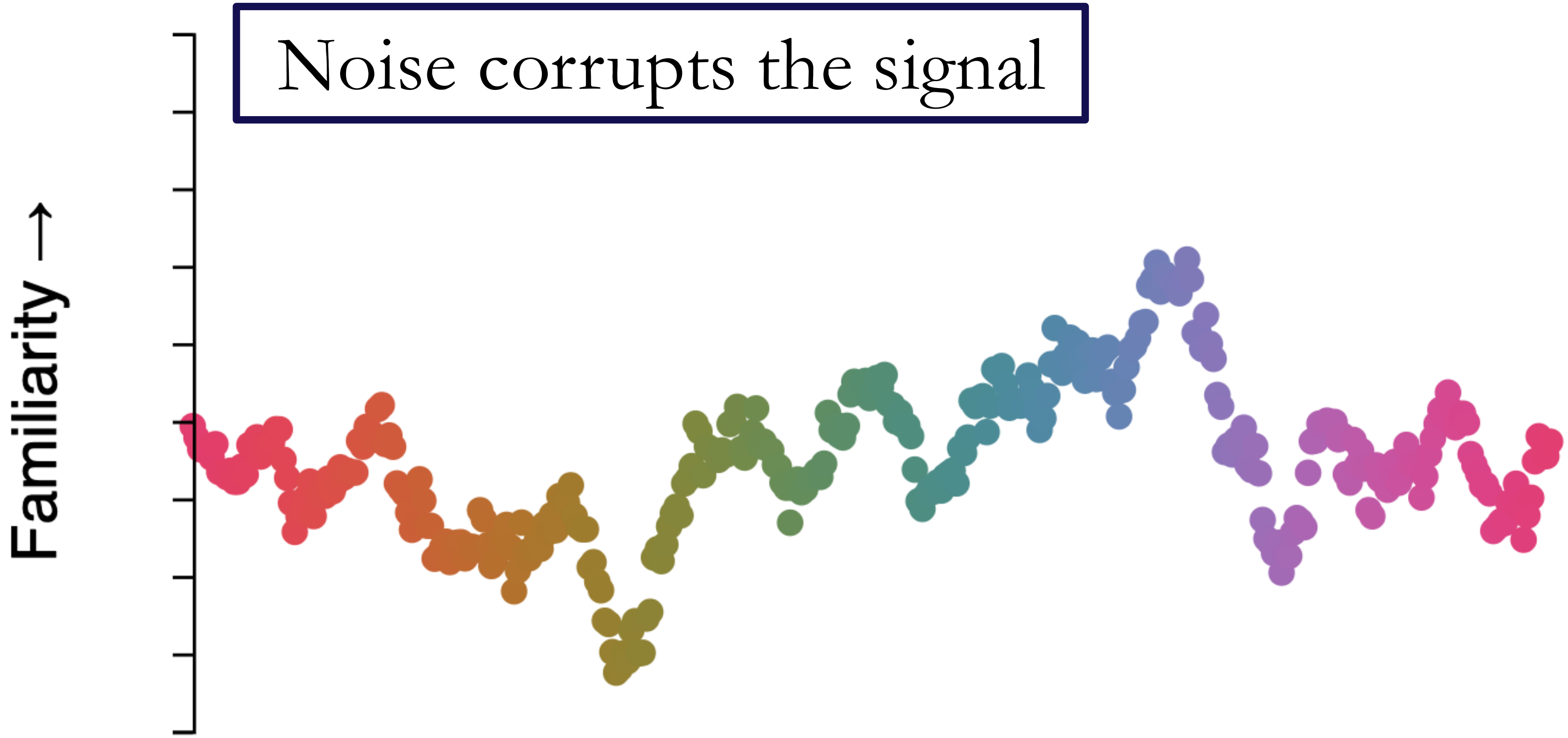
Connecting concepts to working memory



Memory is scaled by psychological similarity.

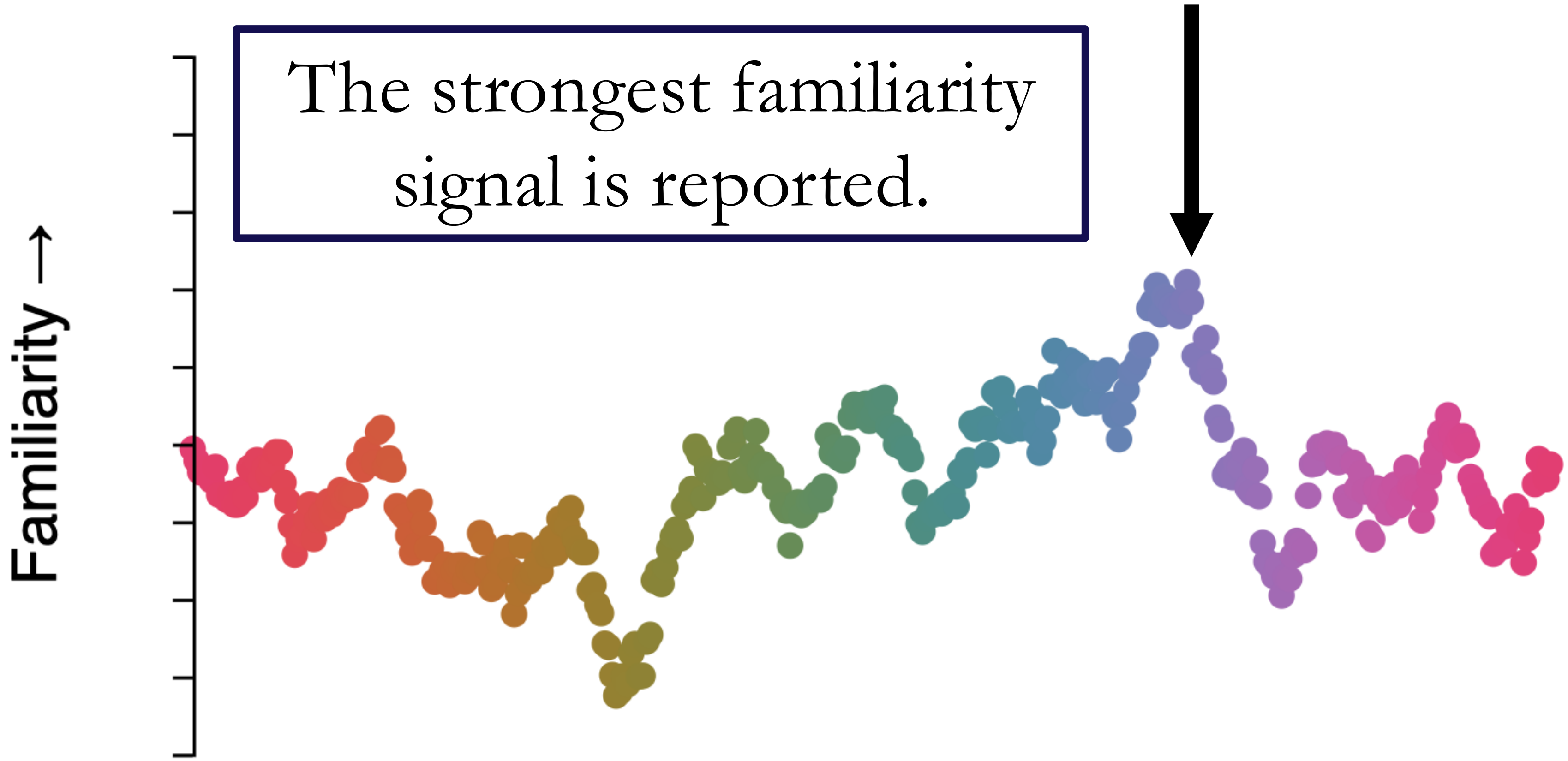
Schurgin, Wixted and Brady (2020) https://doi.org/10.1038/s41562-020-00938-0

Connecting concepts to working memory



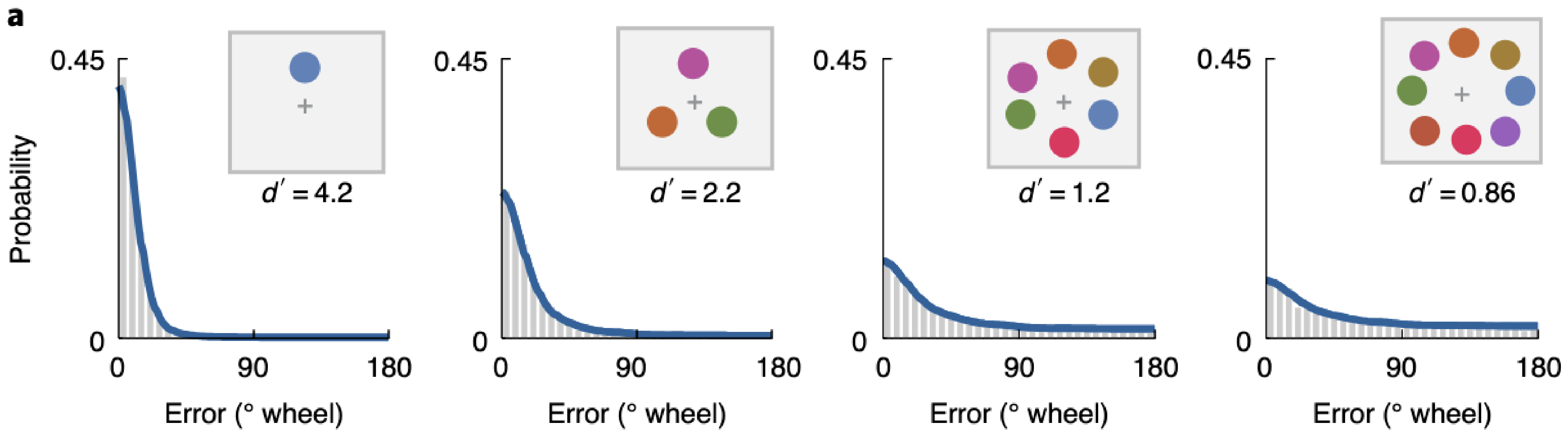
Schurgin, Wixted and Brady (2020) https://doi.org/10.1038/s41562-020-00938-0

Connecting concepts to working memory



Schurgin, Wixted and Brady (2020) https://doi.org/10.1038/s41562-020-00938-0

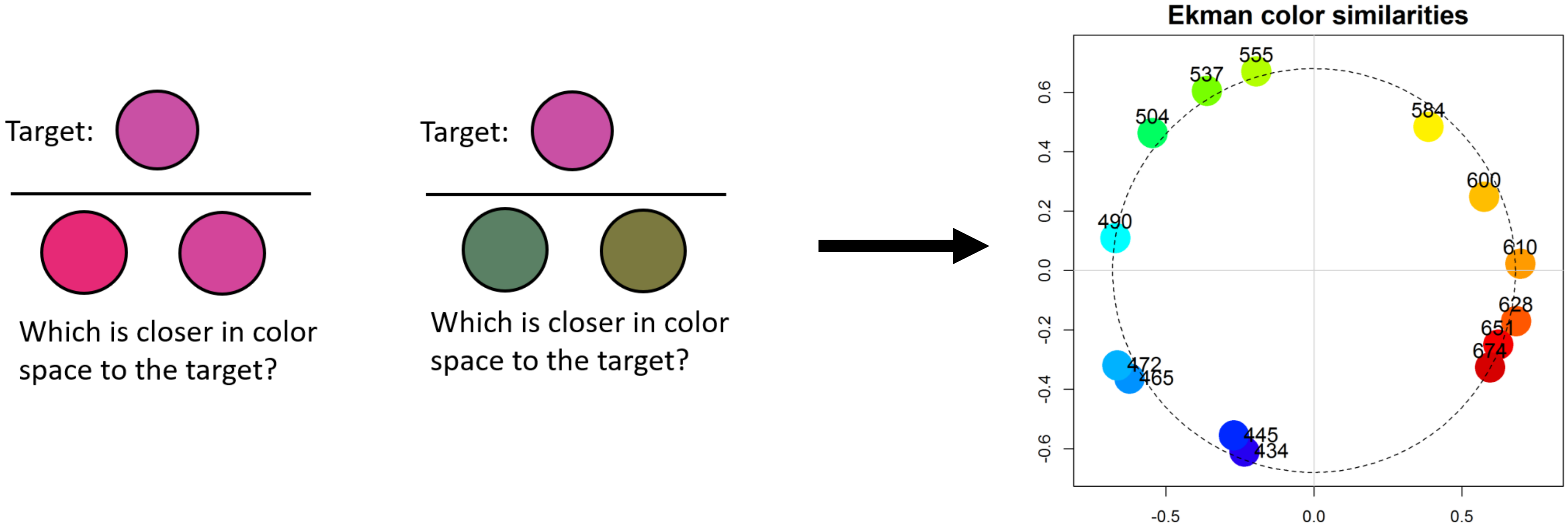
Connecting concepts to working memory



Schurgin, Wixted and Brady (2020) https://doi.org/10.1038/s41562-020-00938-0

The representation in visual working memory

NB. The TCC does not use MDS to project distances; it is a gradient applied across the colour wheel.



Ekman (1954). https://doi.org/10.1080/00223980.1954.9712953
Schurgin, Wixted and Brady (2020) https://doi.org/10.1038/s41562-020-00938-0

See Oberauer, K. (2023). Measurement models for visual working memory—A factorial model comparison. *Psychological review*, *130*(3), 841.

The representation in visual working memory

Recent work has found that adjustments for psychological similarity do not improve model fits for visual working memory performance.

But there remain issues with assuming the physical stimulus space or the similarity-based space to build cognitive models.

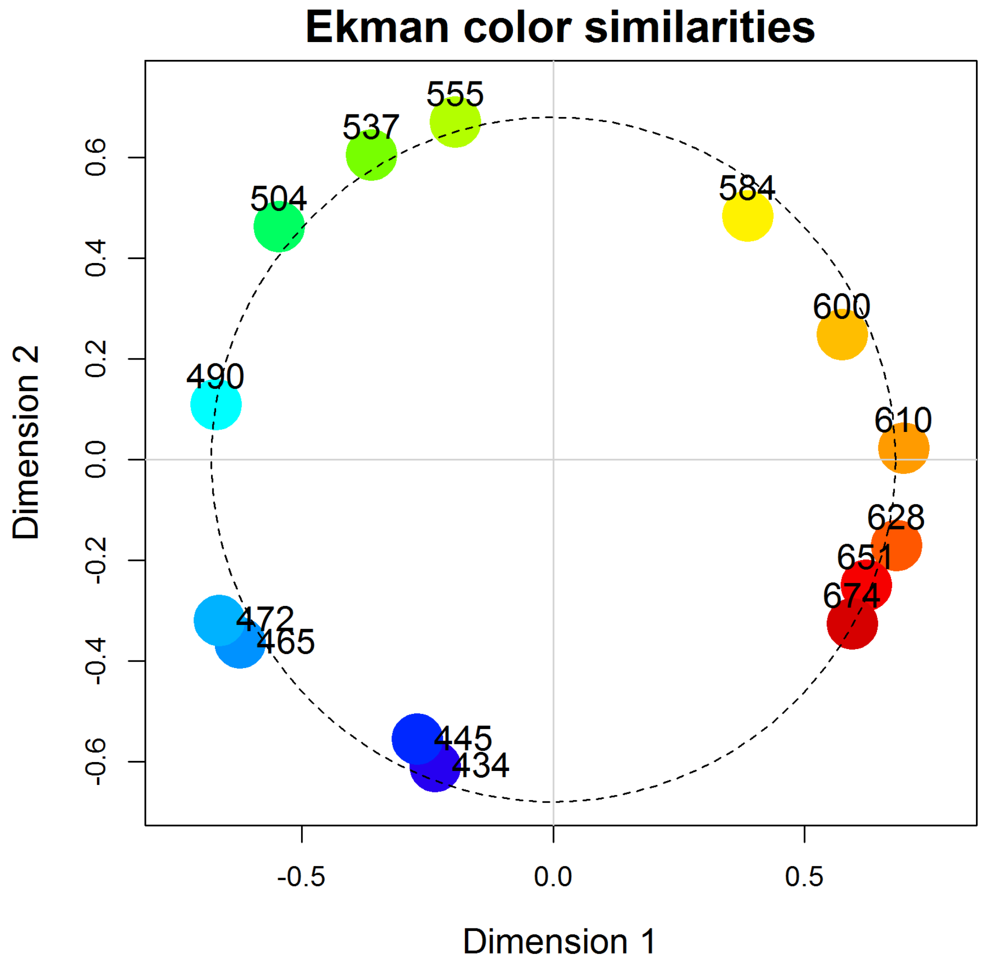
Thus, the question remains – **what is the representation underlying visual working memory?**

Oberauer, K. (2023). Measurement models for visual working memory—A factorial model comparison. *Psychological review*, *130*(3), 841.
Tomić, I., & Bays, P. M. (2024). Perceptual similarity judgments do not predict the distribution of errors in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 50(4), 535.

The representation in visual working memory

Can the **representational geometry** predict the error distributions in working memory tasks? (see Wei and Woodford, 2025)

Does psychological similarity and working memory share the same representation?



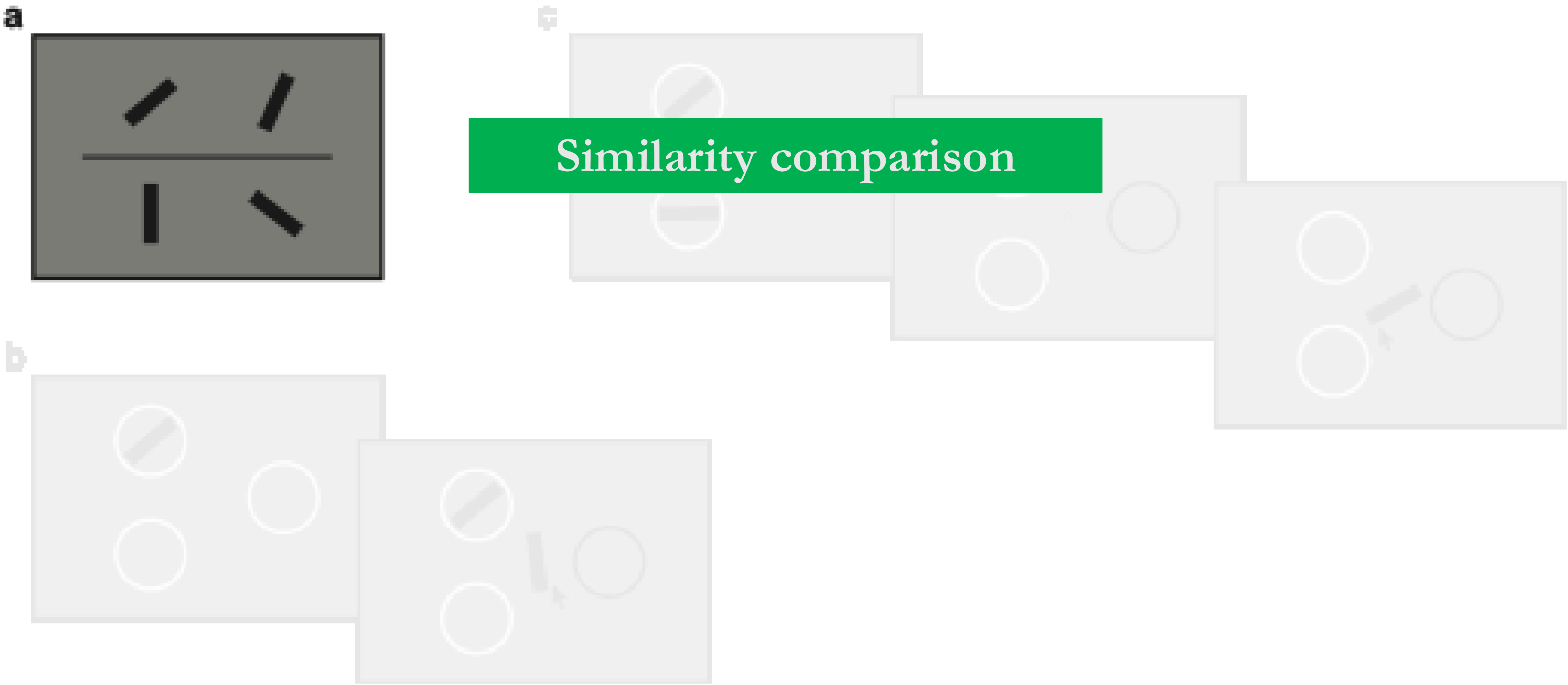
Ekman, G. (1954). Dimensions of color vision. *The Journal of Psychology*, *38*(2), 467-474.
Wei, X. X., & Woodford, M. (2025). Representational geometry explains puzzling error distributions in behavioral tasks. Proceedings of the National Academy of Sciences, 122(4), e2407540122.

Our modeling approach

In brief, we used Bayesian MCMC methods to recover the **latent representation** underlying three cognitive tasks using oriented lines. The data comes from an open dataset (Tomic and Bays, 2024).

Tomić, I., & Bays, P. M. (2024). Perceptual similarity judgments do not predict the distribution of errors in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *50*(4), 535.

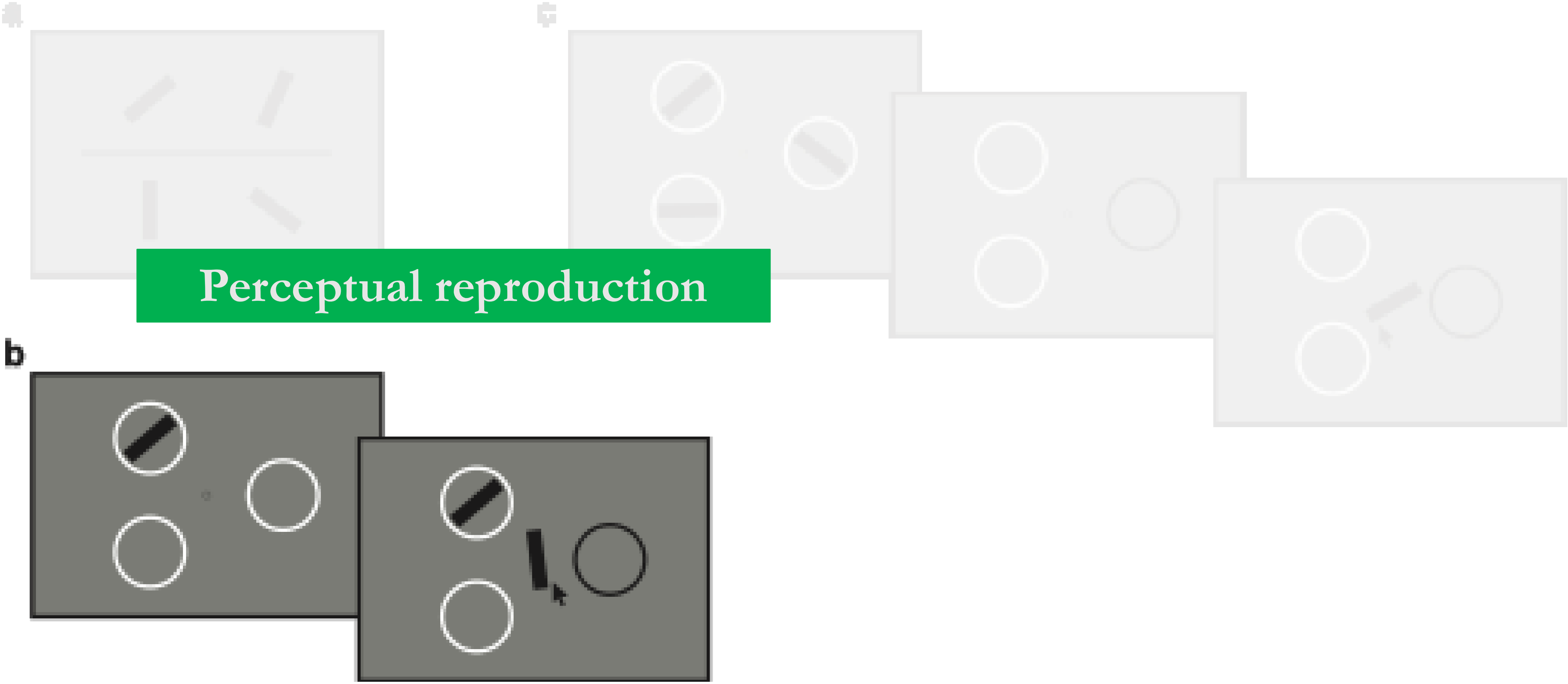
The cognitive tasks



Similarity comparison

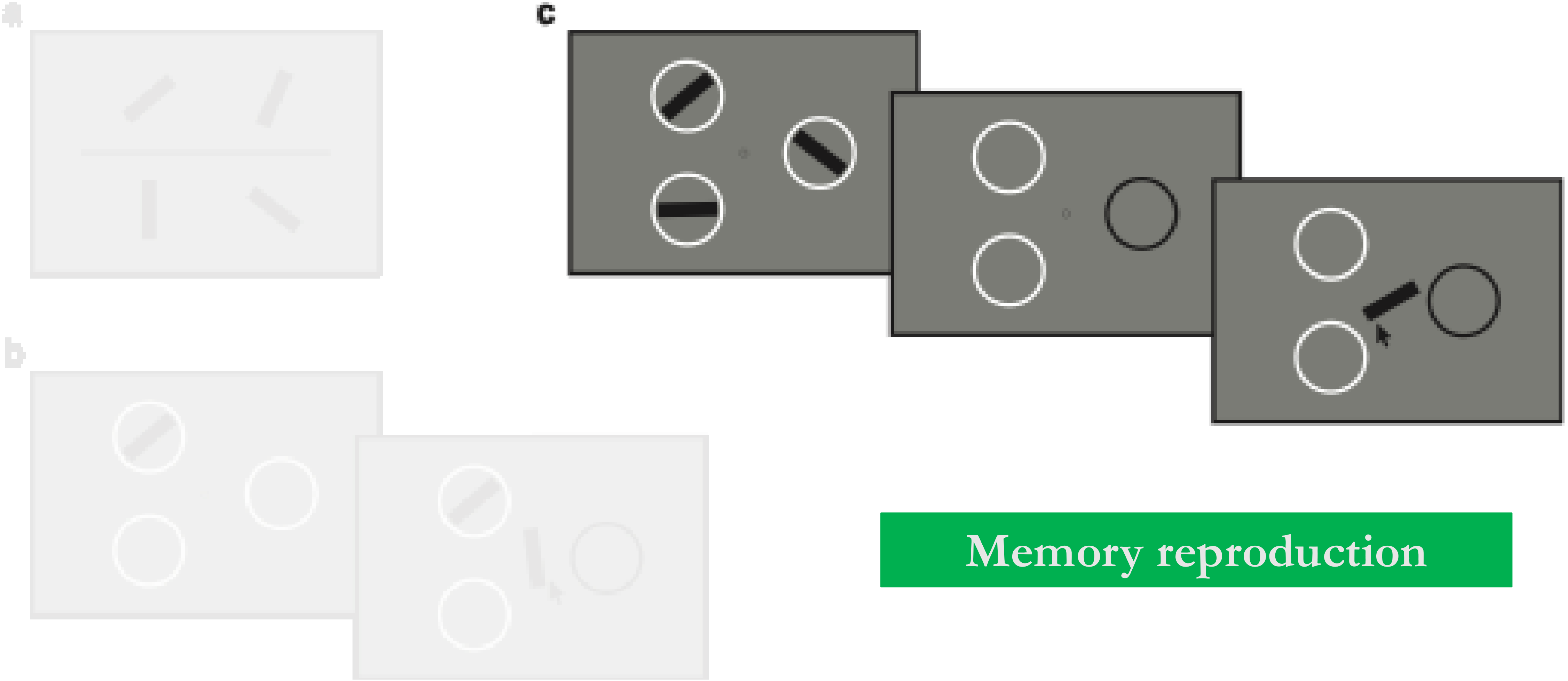
Tomić, I., & Bays, P. M. (2024). Perceptual similarity judgments do not predict the distribution of errors in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 50(4), 535.

The cognitive tasks



Tomić, I., & Bays, P. M. (2024). Perceptual similarity judgments do not predict the distribution of errors in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 50(4), 535.

The cognitive tasks



Memory reproduction

Tomić, I., & Bays, P. M. (2024). Perceptual similarity judgments do not predict the distribution of errors in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 50(4), 535.

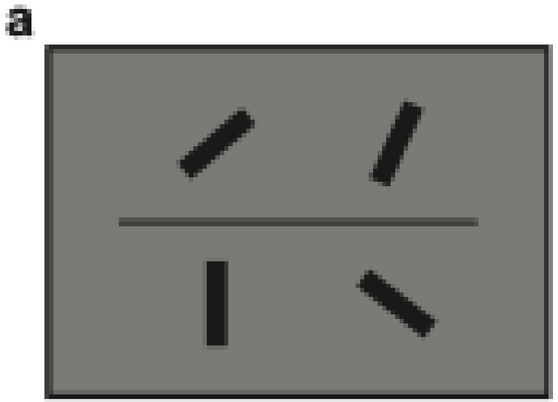
Our modeling goal

Our goal was to recover the **latent representations** of the orientation stimuli in the three cognitive tasks.

Note that the latent representation is **modelled using the task data itself**, not outsourced to MDS with similarity judgments.

Our modeling approach

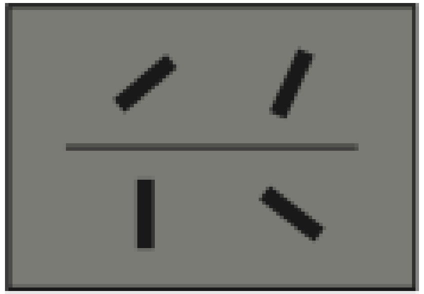
We took a Bayesian generative modeling approach – we built a **Thurstonian model** that infers the underlying representation.

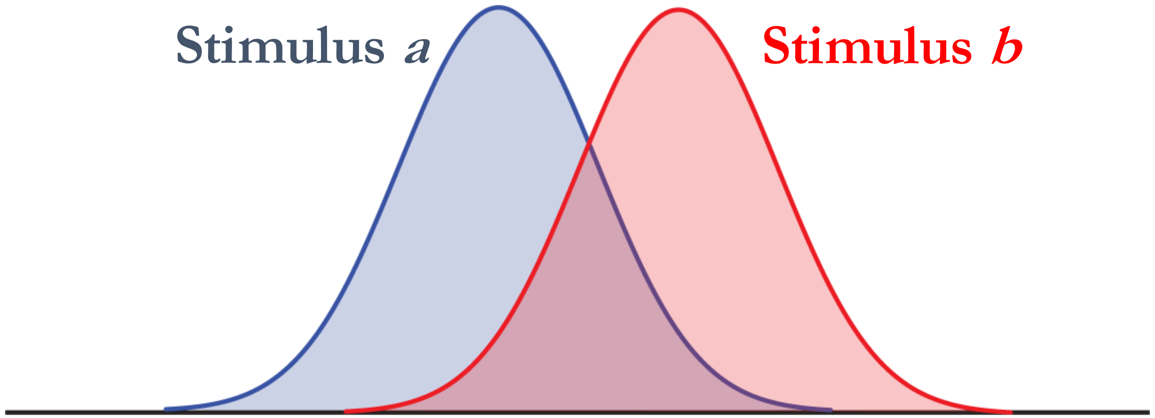


Similarity comparison

Our modeling approach

a



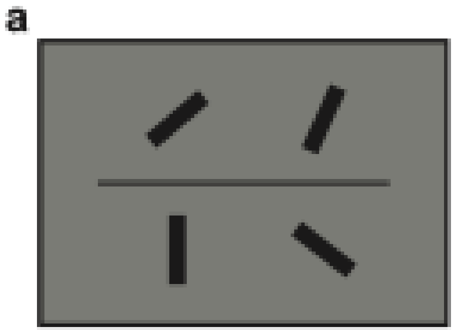


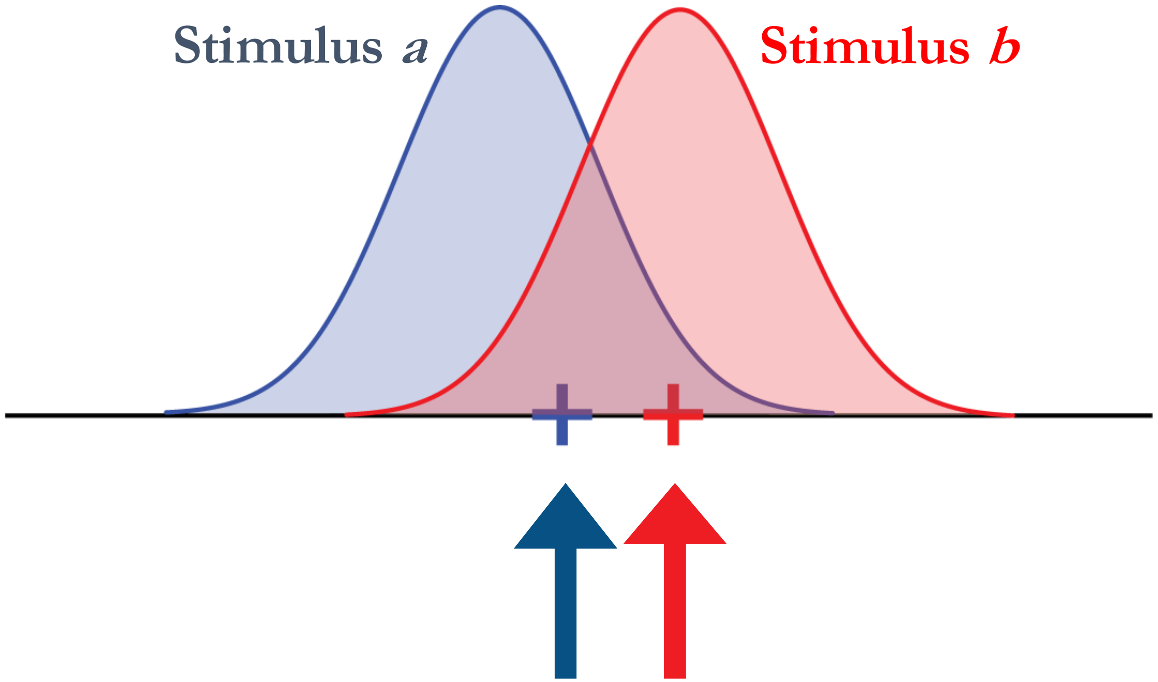
We model the latent representation of each orientation stimulus as a circular Gaussian distribution, centred on a value (μ_n) and with a common width (σ).

For model identifiability $\begin{cases} \mu_1 = 0 \\ \mu_2 \sim \text{uniform}(0\,, \frac{\pi}{2}) \end{cases}$

Uninformed priors $\begin{cases} \mu_3, \ldots \mu_n \sim \text{uniform}(0\,, \pi) \\ \sigma \sim \text{uniform}(0\,, \pi) \end{cases}$

Our modeling approach





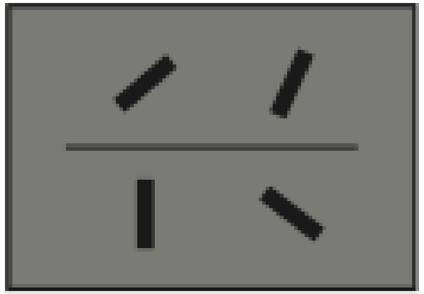
The participant takes one sample from each latent representation of the pair.

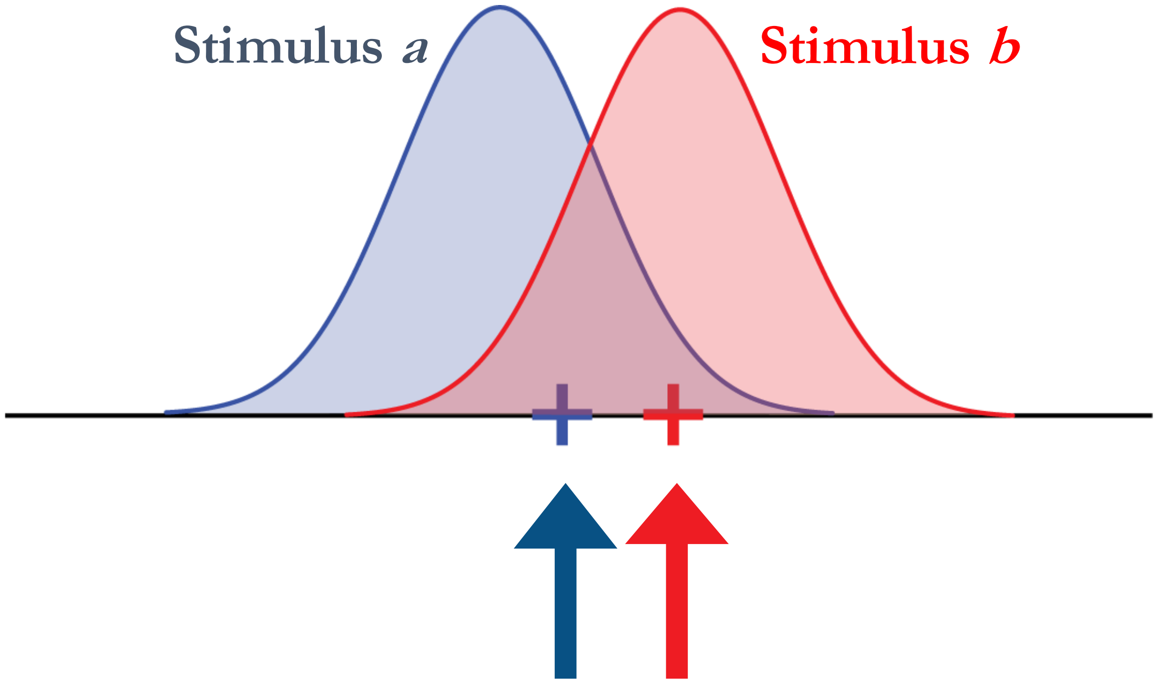
$$x_t^a \sim \text{Gaussian}(\mu_{a_t}, \frac{1}{\sigma^2})$$

$$x_t^b \sim \text{Gaussian}(\mu_{b_t}, \frac{1}{\sigma^2})$$

Our modeling approach

a



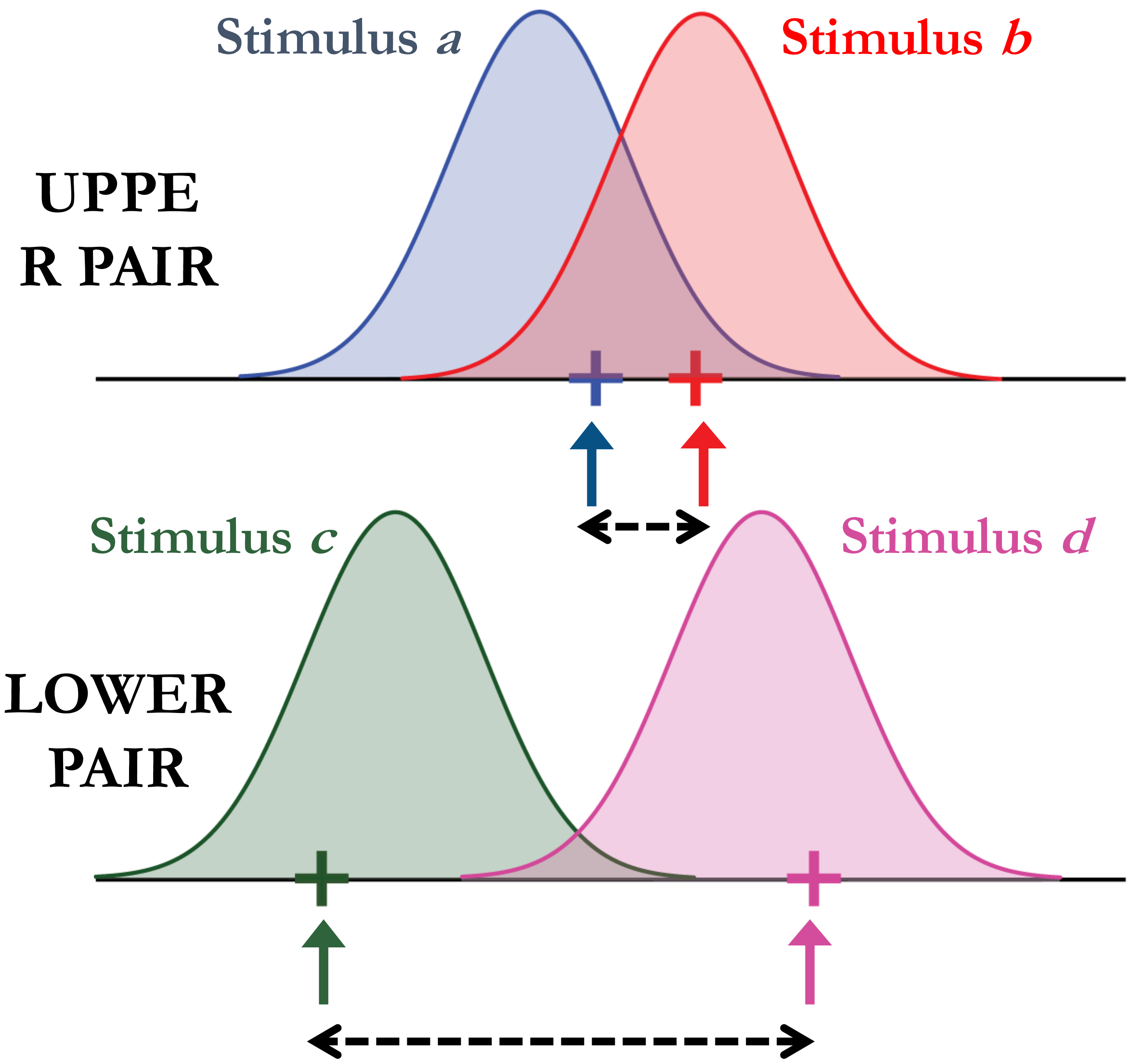


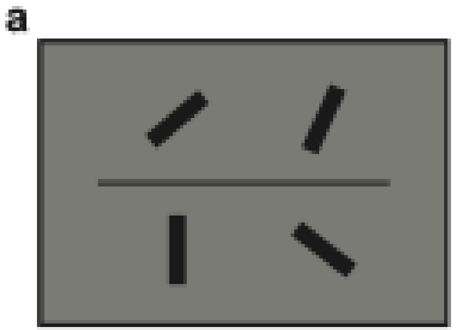
The psychological distance between the pair of stimuli is the distance between the samples

$$d_t^{ab} = \min(|x_t^a - x_t^b|, \pi - |x_t^a - x_t^b|)$$

Accounts for semi-circular nature of angles

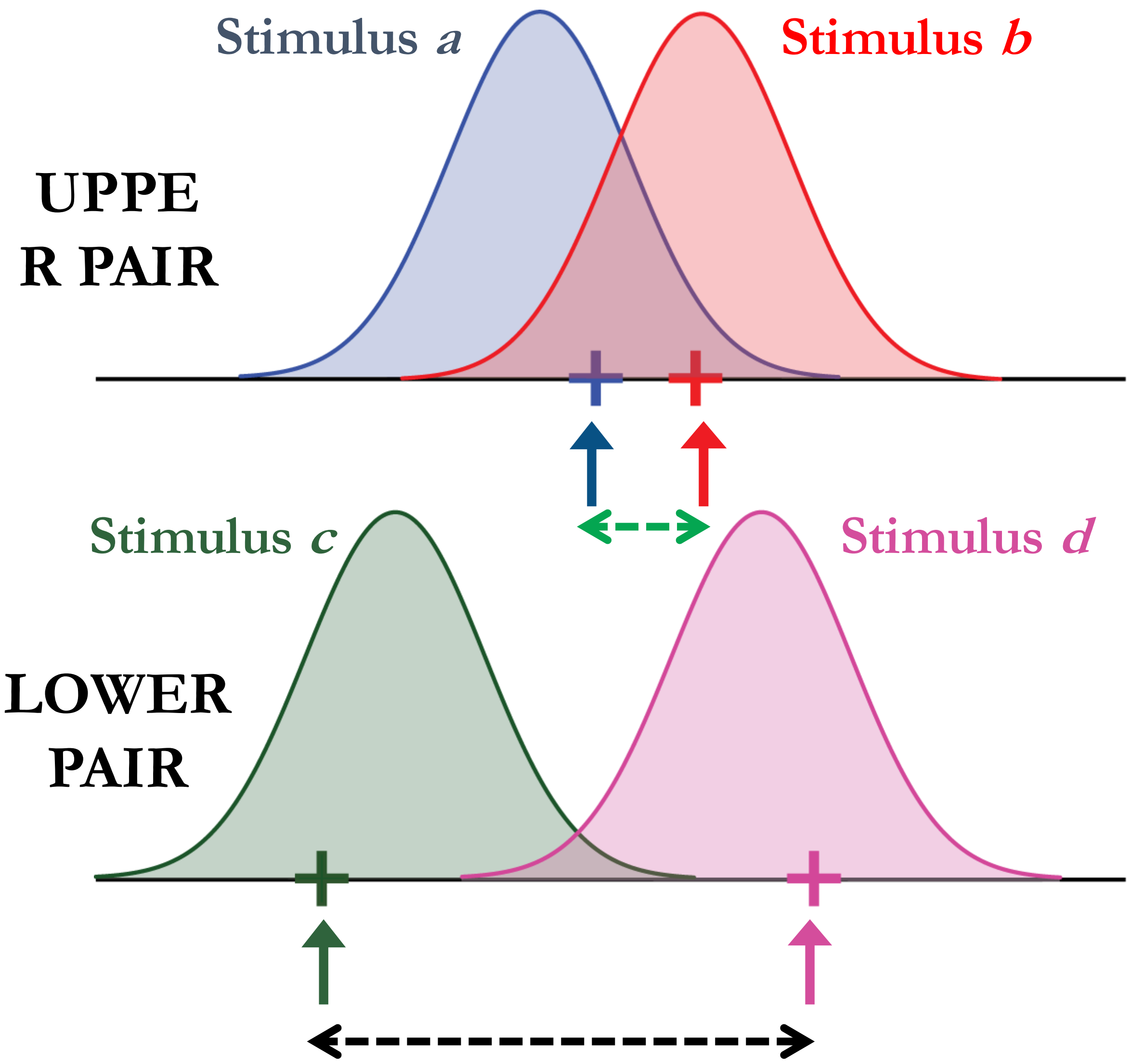
Our modeling approach

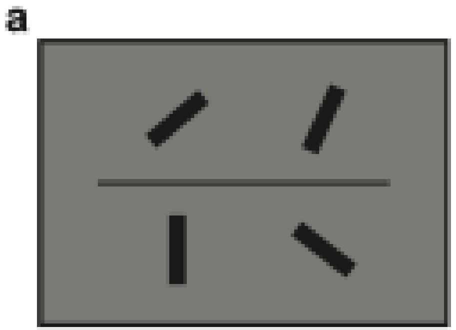




Repeat to calculate the psychological distance for the orientation stimuli in the second pair.

Our modeling approach

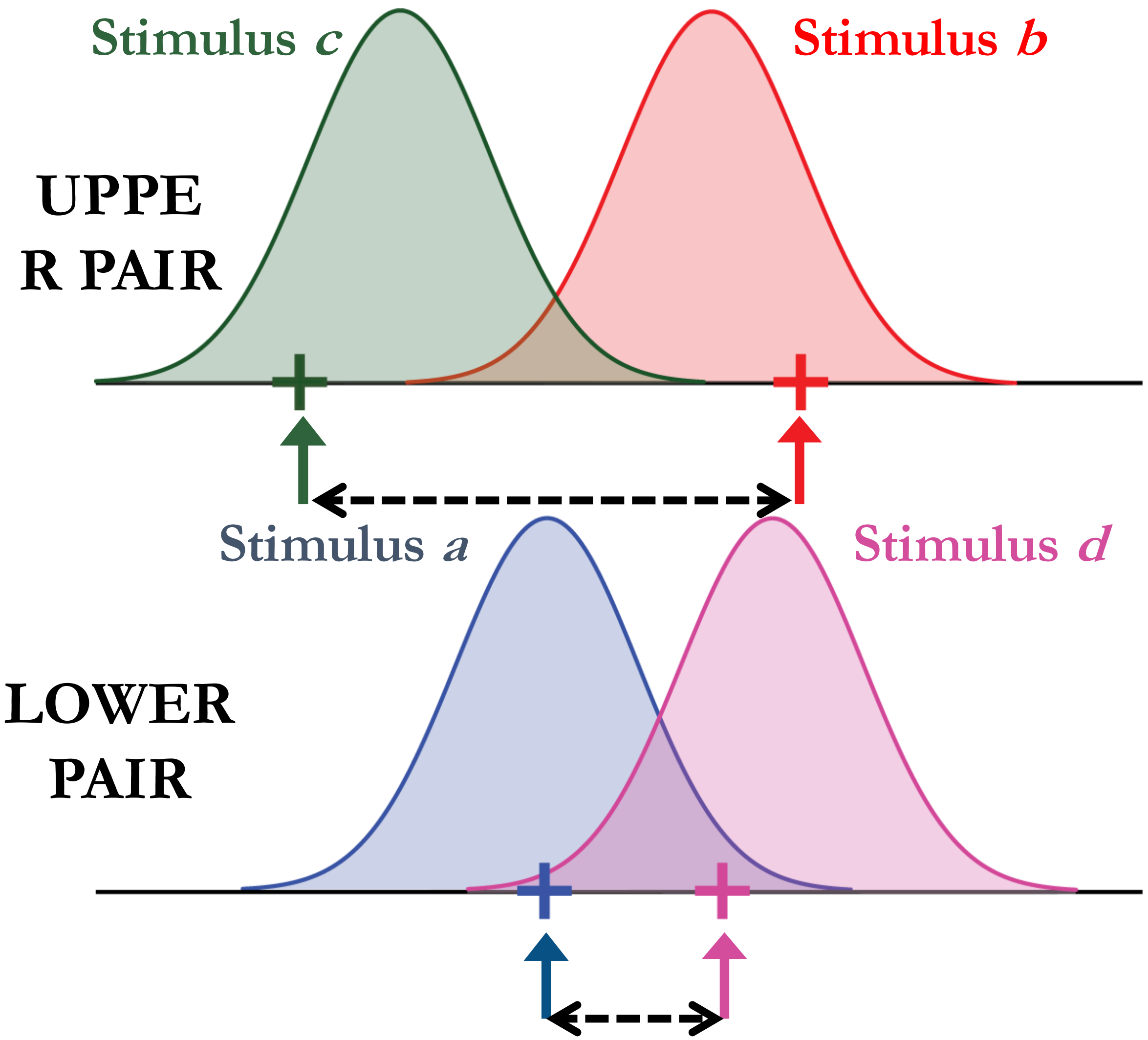


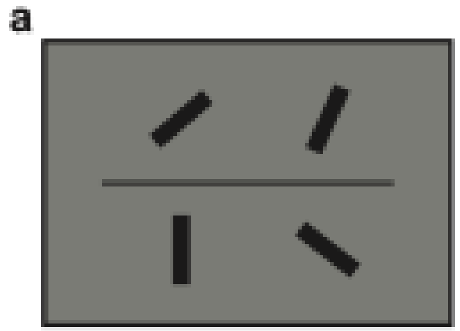


Repeat to calculate the psychological distance for the orientation stimuli in the second pair.

The shorter distance determines which pair is selected as being **more similar.**

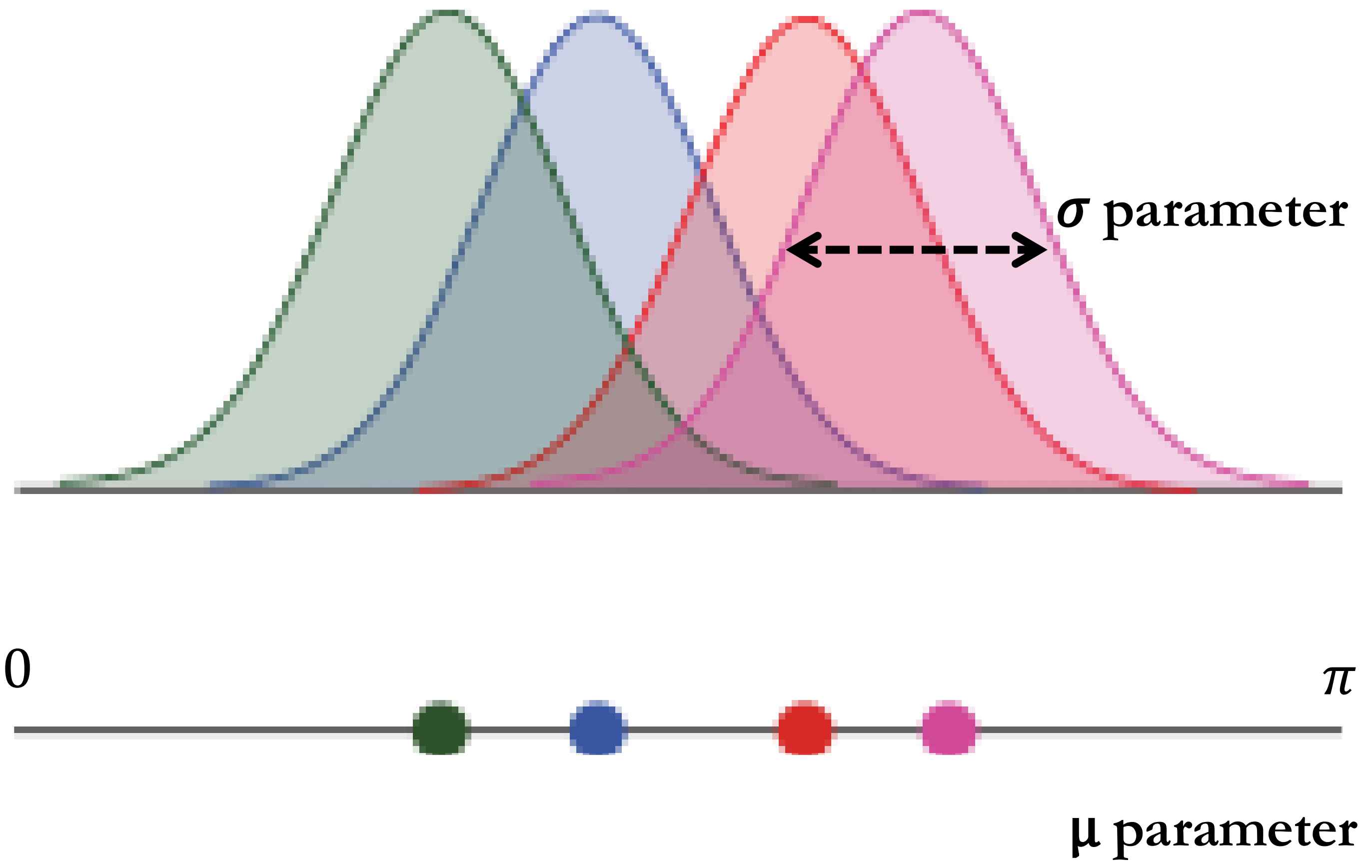
Our modeling approach





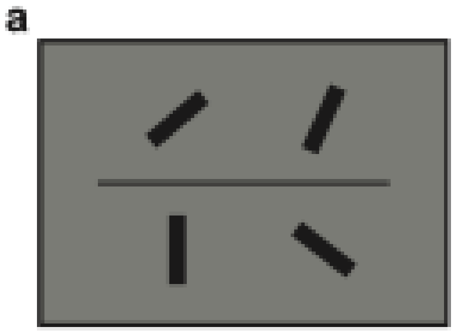
Repeat over a number of trials.

Our modeling approach

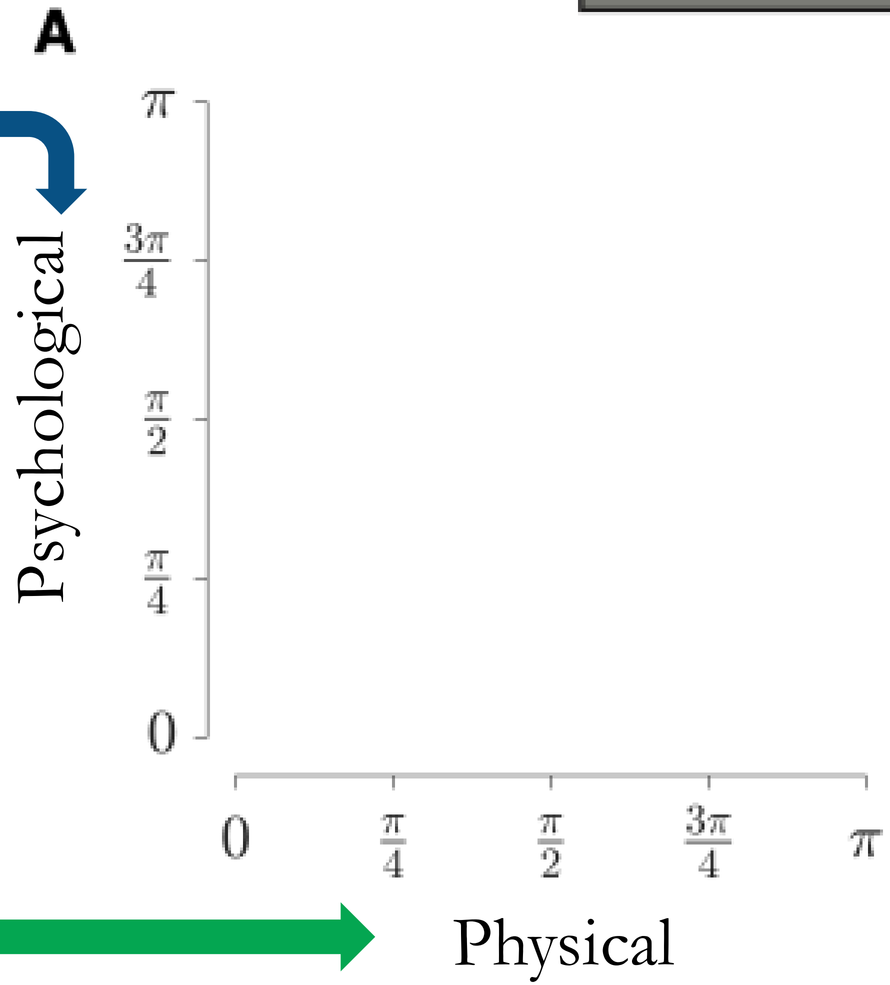


The model infers the latent representations – the center of the distribution, and their common width – based on the response data.

Similarity comparison



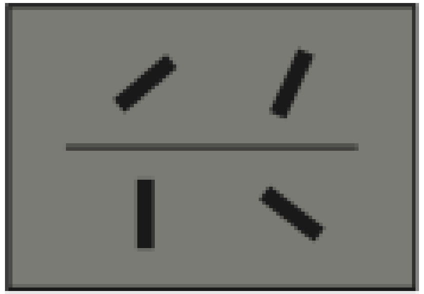
The mental representation of the line orientation



The orientation angle of the presented line stimulus

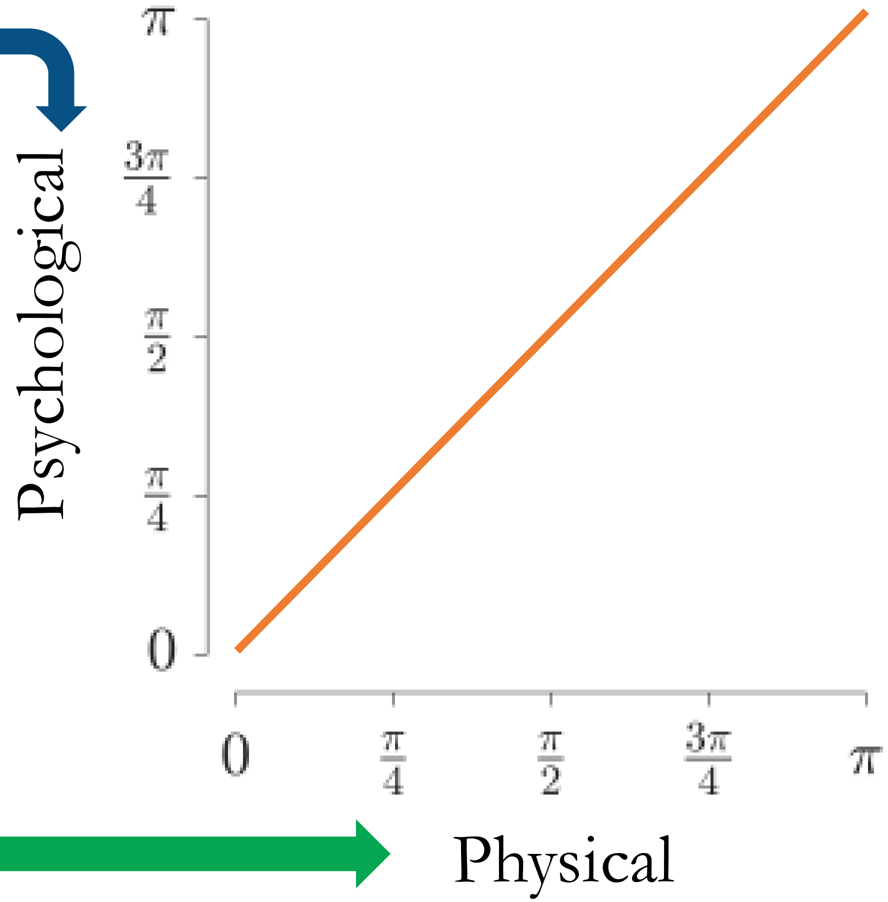
Similarity comparison

a



A

The mental representation of the line orientation

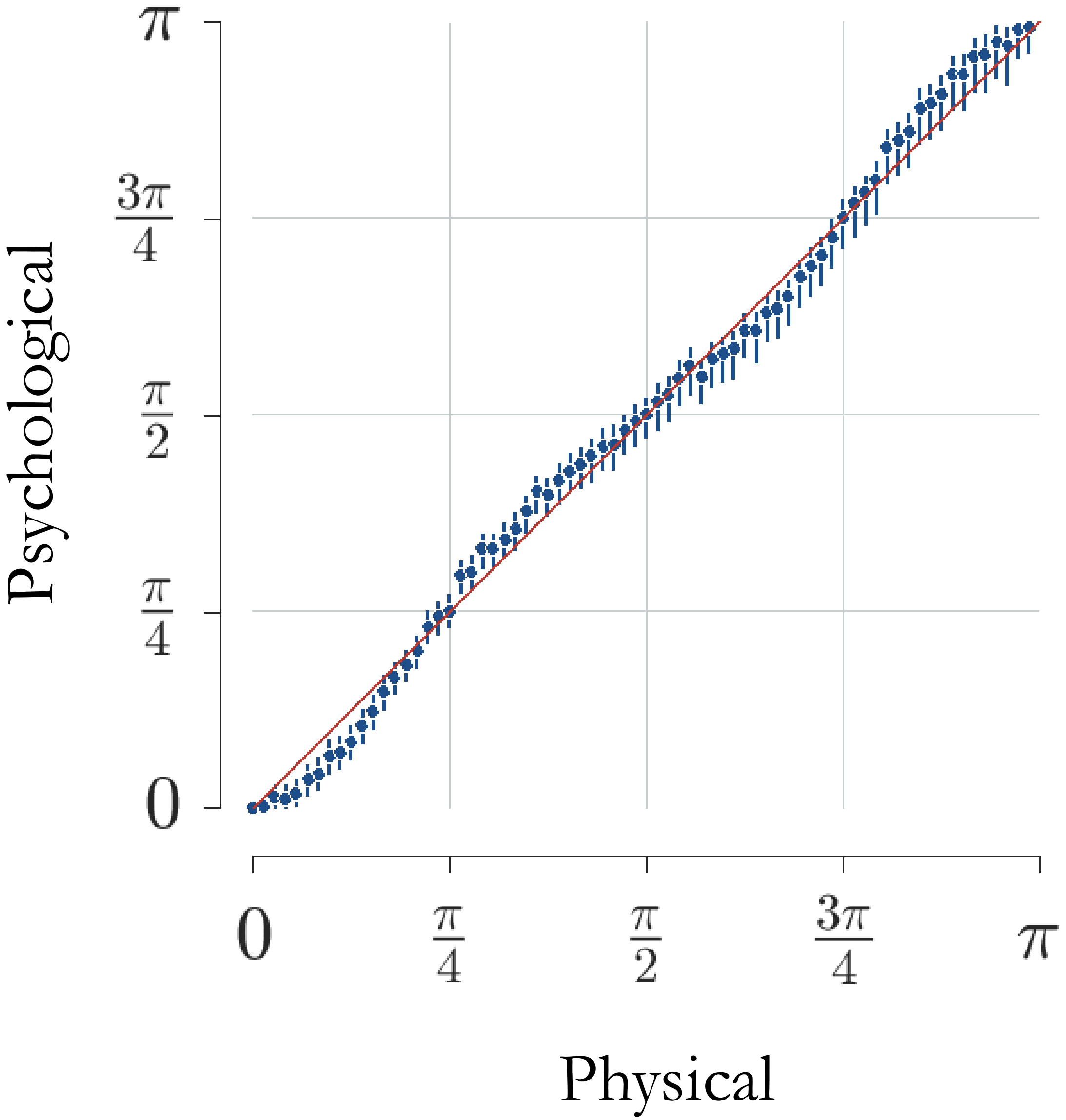


The orientation angle of the presented line stimulus

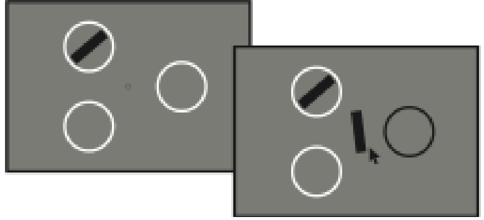
Similarity comparison

The representation **does not match the physical stimulus space** – it is not exactly a diagonal line.

Clear deviations where close to vertical lines appear more vertical, and close to horizontal lines appear more horizontal.

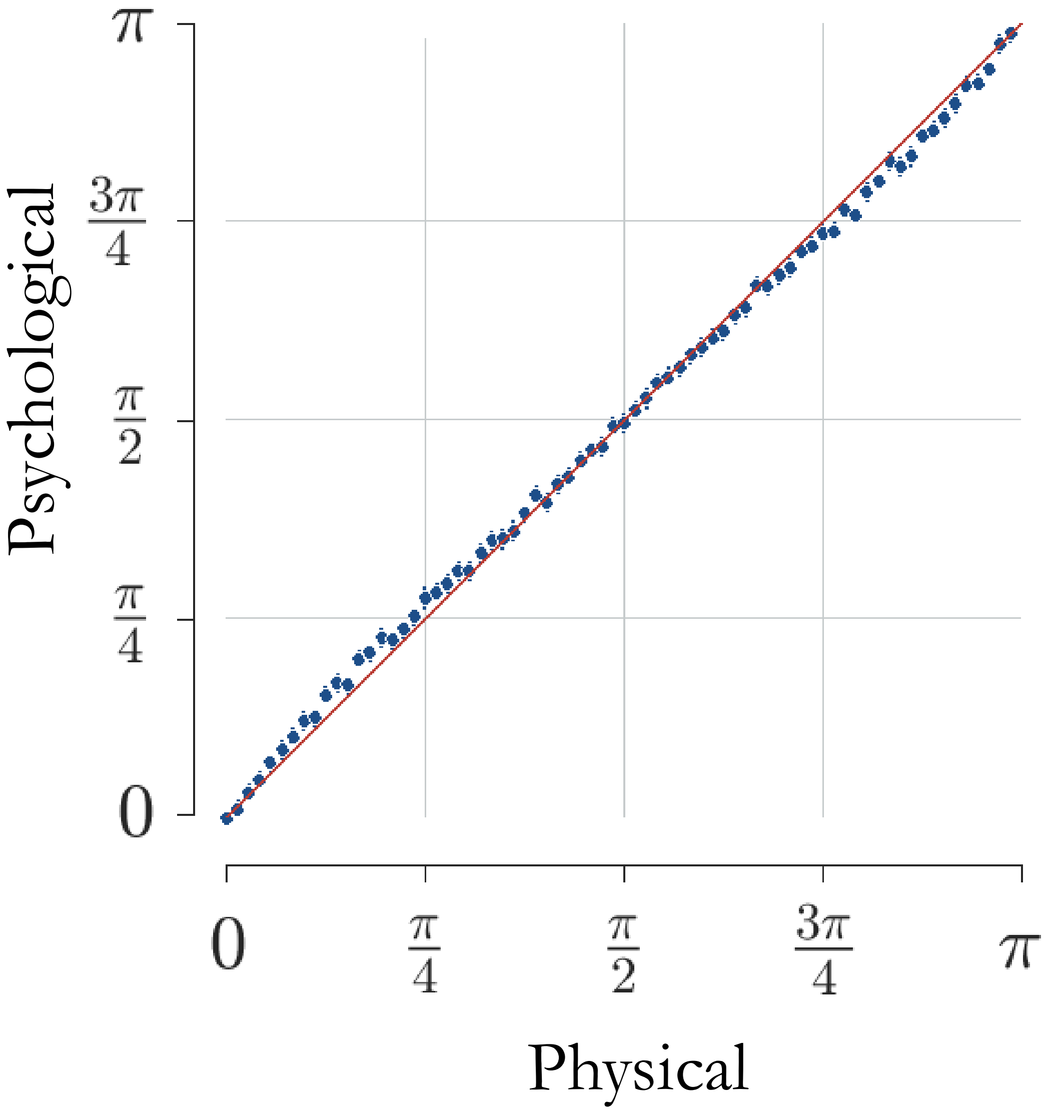


Perceptual reproduction



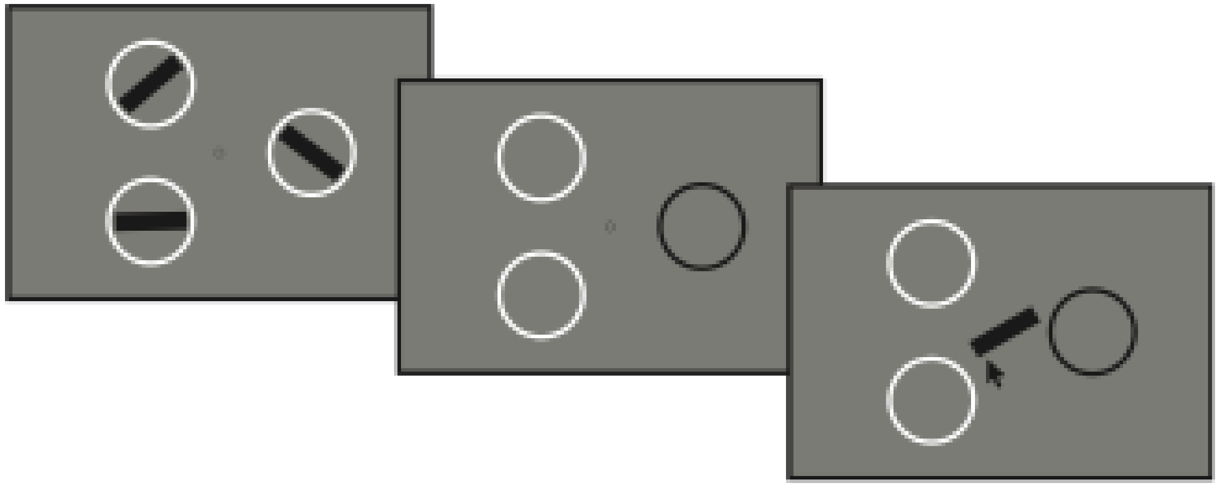
The representation is very close to the physical stimulus space, but there are slight deviations.

Not quite horizontal or not quite vertical lines are slightly biased away from the cardinal directions. This is the opposite to the similarity comparison.



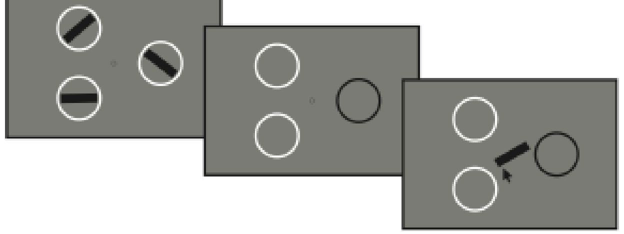
Memory reproduction

NB. The memory reproduction model includes a **swap error** process, where a sample is drawn from the latent representation of another item in the array.

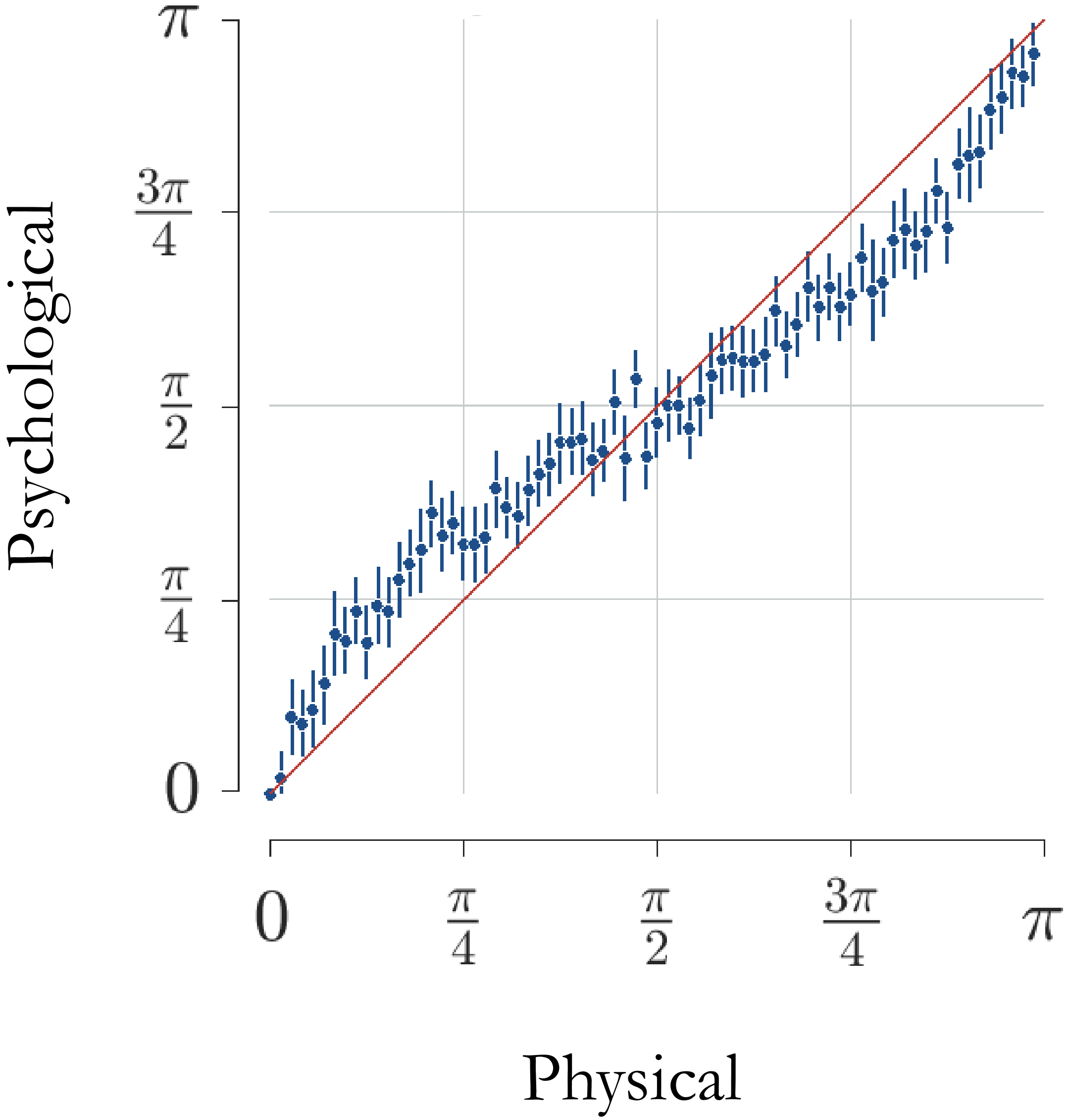


Bays, P. M. (2016). Evaluating and excluding swap errors in analogue tests of working memory. *Scientific reports*, *6*(1), 19203.

Memory reproduction



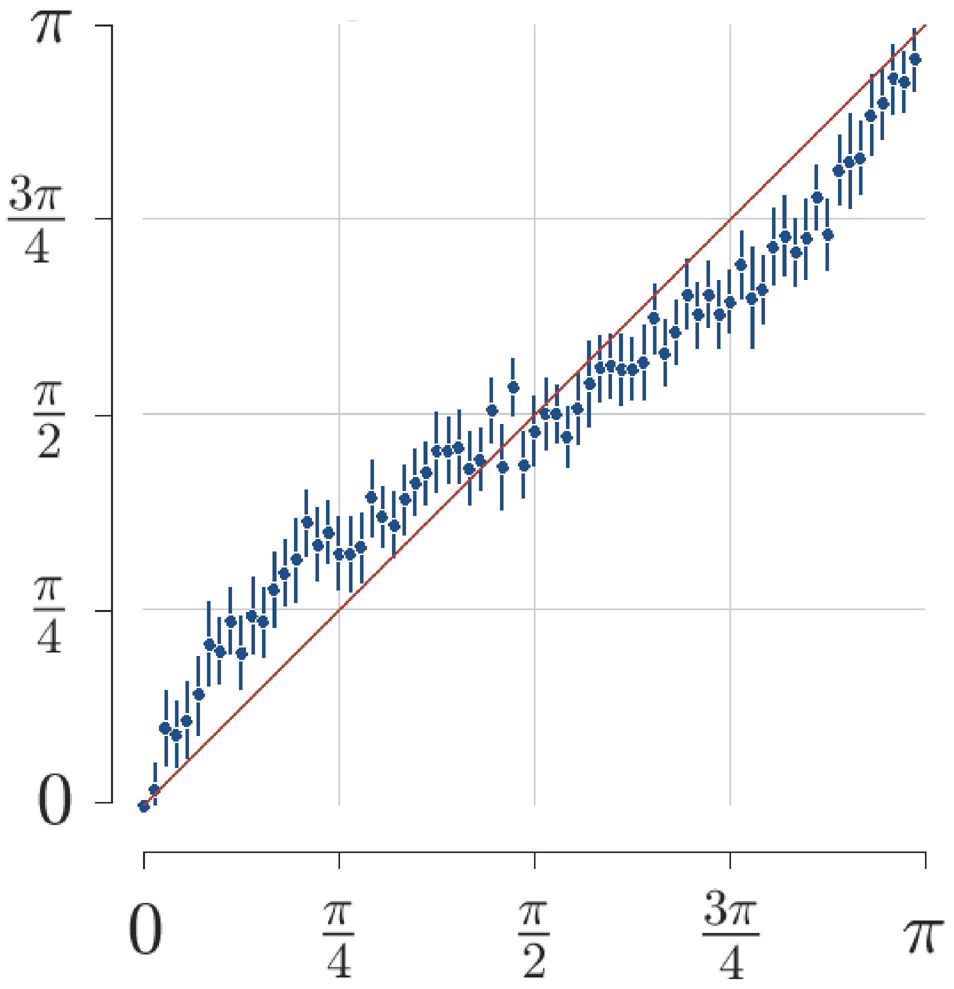
The representation is similar to the perceptual reproduction task but with larger deviations.



Lines are represented more towards the oblique directions than they actually are. This is the **opposite to similarity comparison**.

Conclusions

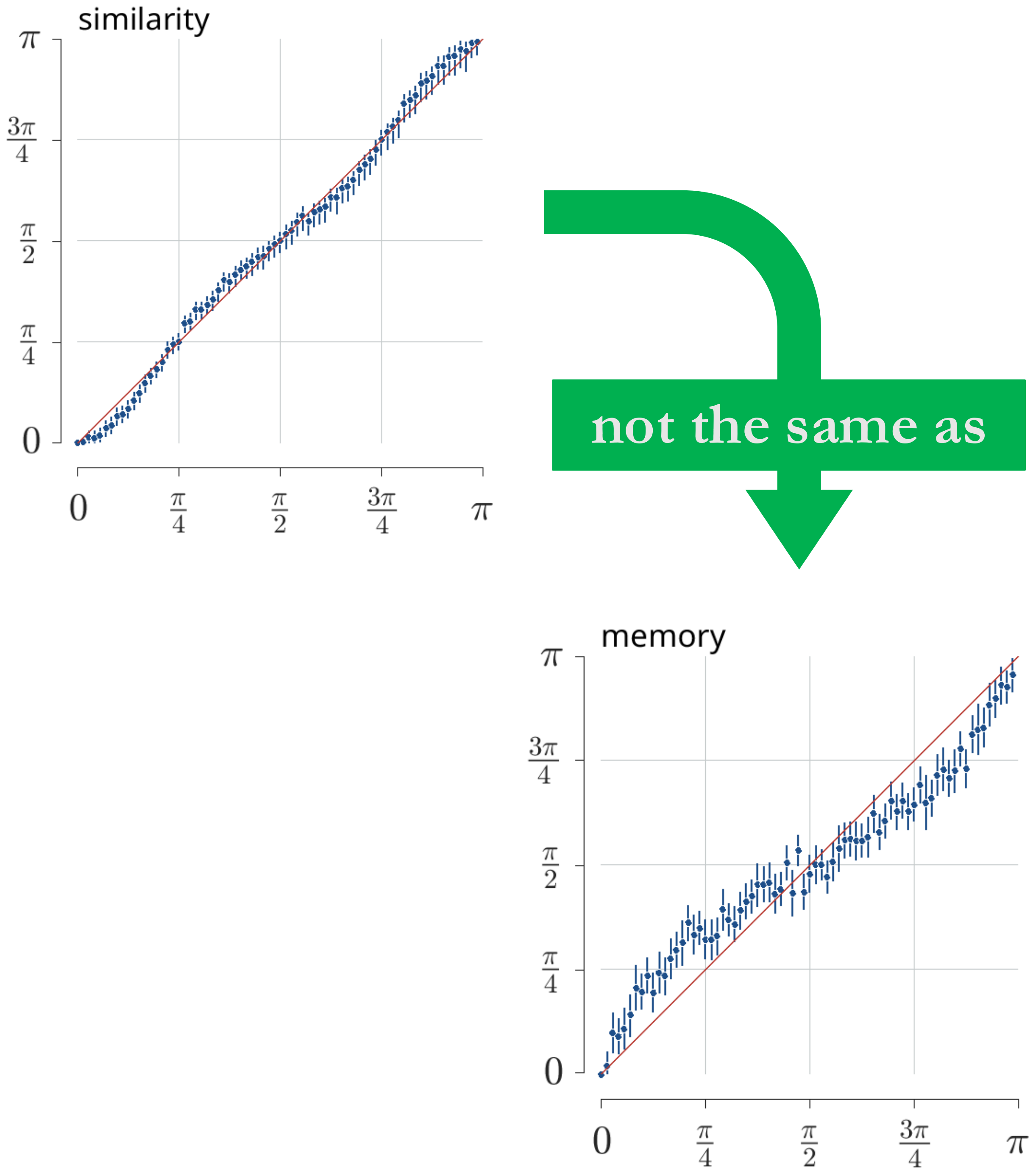
The cognitive representation for memory reproduction does not match the physical stimulus space. Therefore, models **should not assume the physical stimulus space** when modelling working memory performance.



Conclusions

Similarity comparisons and both reproduction tasks **do not share the same cognitive representation.** (log BF = 495)

Therefore, psychological similarity cannot be assumed to be the basis for working memory.



Conclusions

We have provided preliminary evidence that **representations** are task-specific rather than sharing the same representations.

We advocate for our modeling approach that fits the cognitive representation **jointly** with the theorised mechanisms underlying the tasks.

Conclusions

As we are fitting the representation and processes jointly, changing the processes will change the inferred representation. There are two approaches:

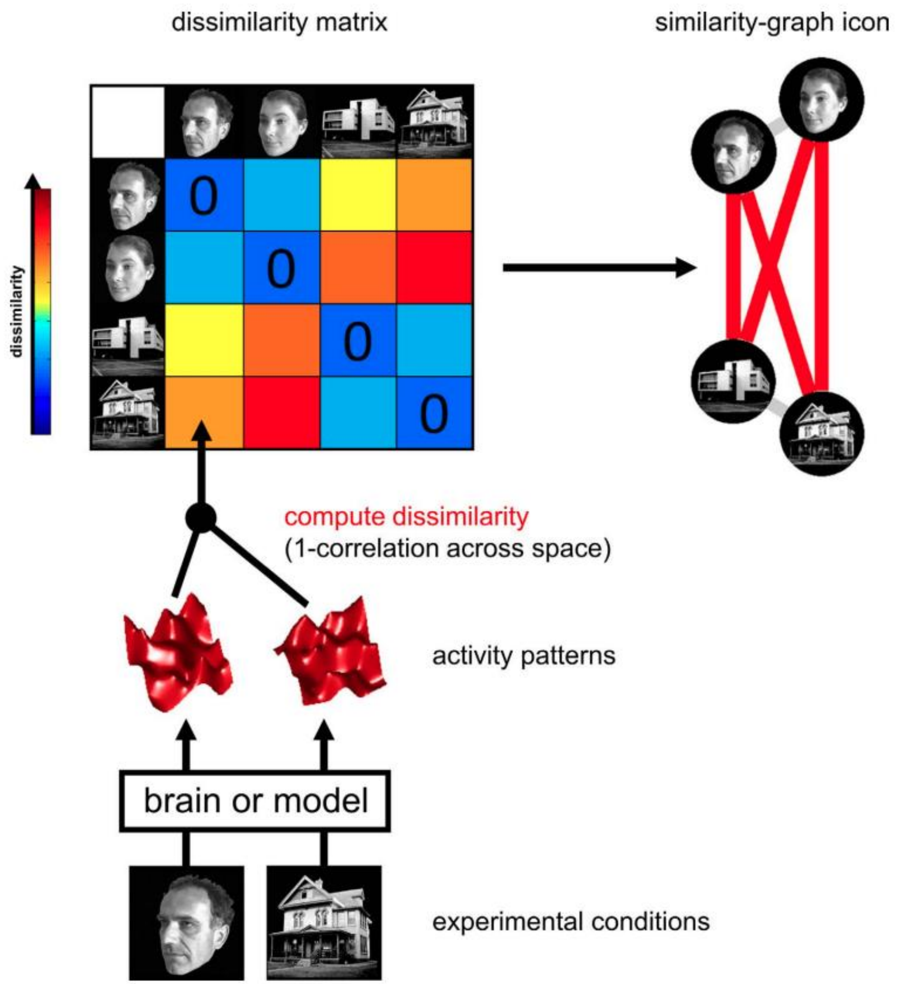
1. **Unifying** the representation by explaining task-specific differences (i.e. direction of bias)

2. **Cataloguing** when the representation is variant or invariant across conditions where the mechanism is believe to be the same

Representations to neuroscientists

Cognitive neuroscientists are training neural network models to achieve human-like performance or using machine learning to decode neural representations.

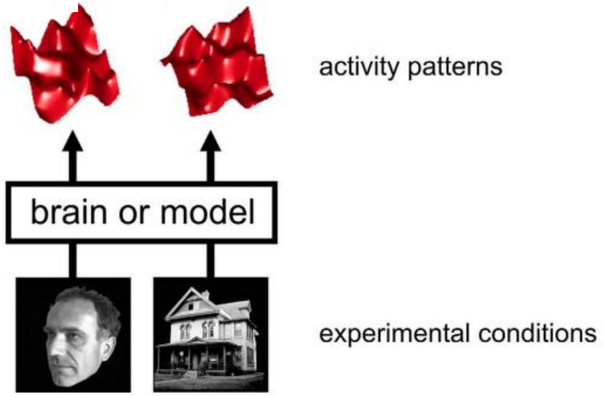
One such method is **representational similarity analysis:**



Kriegeskorte, Mur, & Bandettini (2008). https://doi.org/10.3389/neuro.06.004.2008

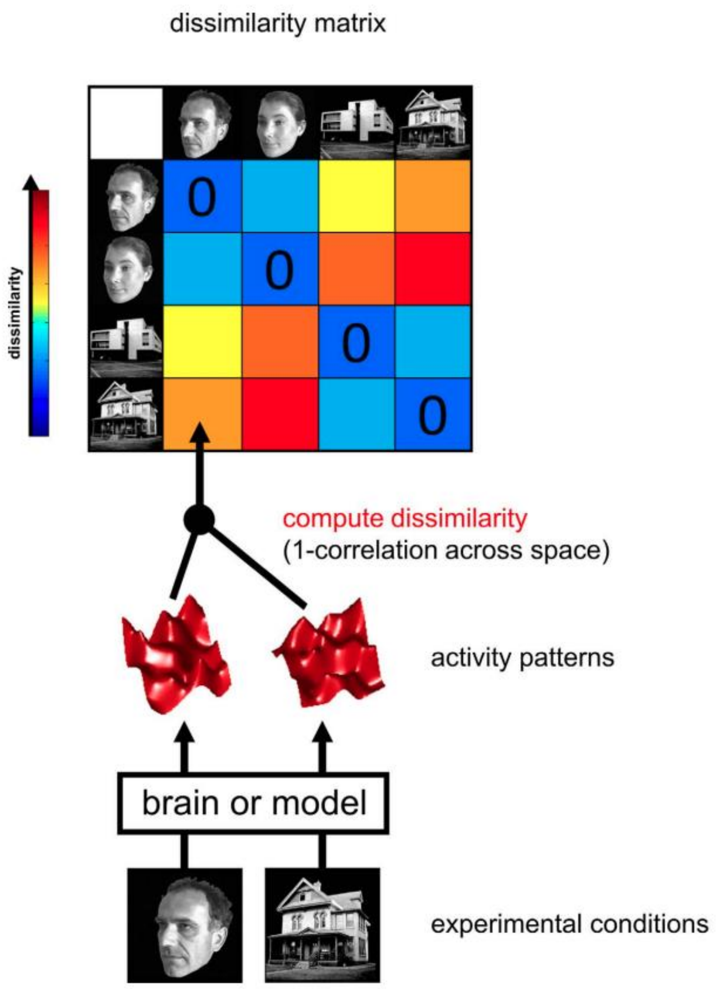
Representational similarity analysis

The researcher collects neuroimaging data under experimental conditions



Kriegeskorte, Mur, & Bandettini (2008). https://doi.org/10.3389/neuro.06.004.2008

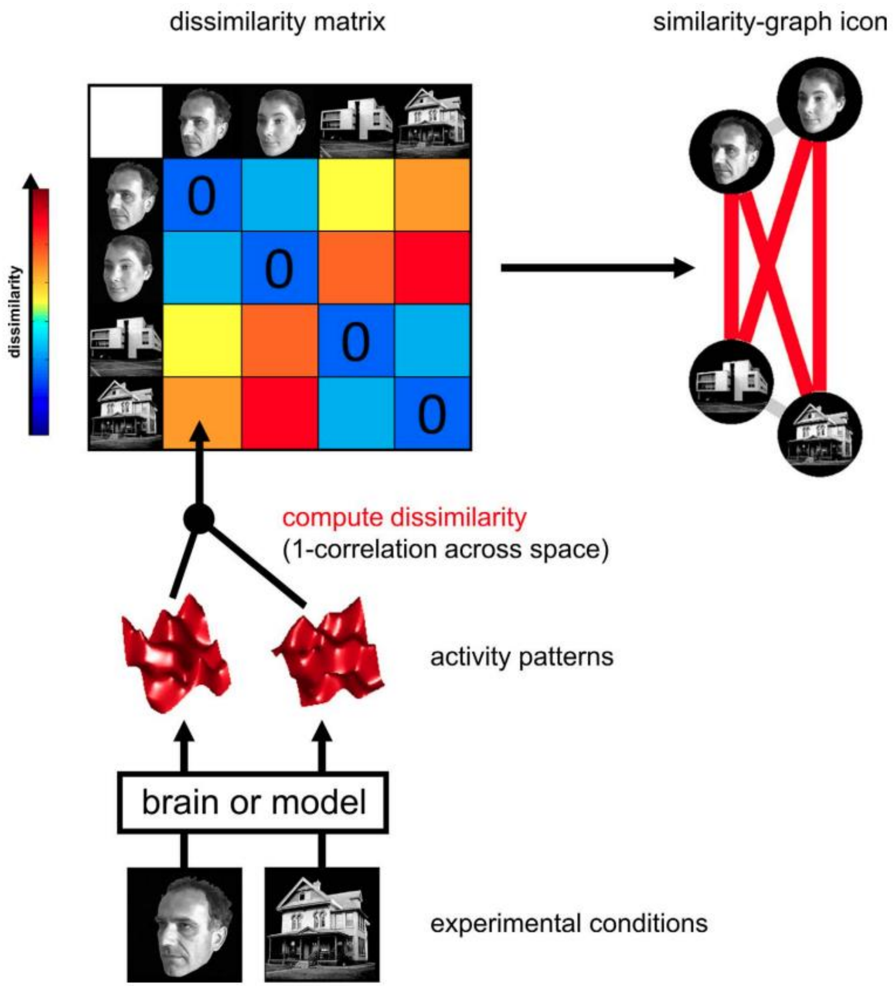
Representational similarity analysis



The researcher uses machine learning to decode the conditions and computes the distance between conditions

Kriegeskorte, Mur, & Bandettini (2008). https://doi.org/10.3389/neuro.06.004.2008

Representational similarity analysis

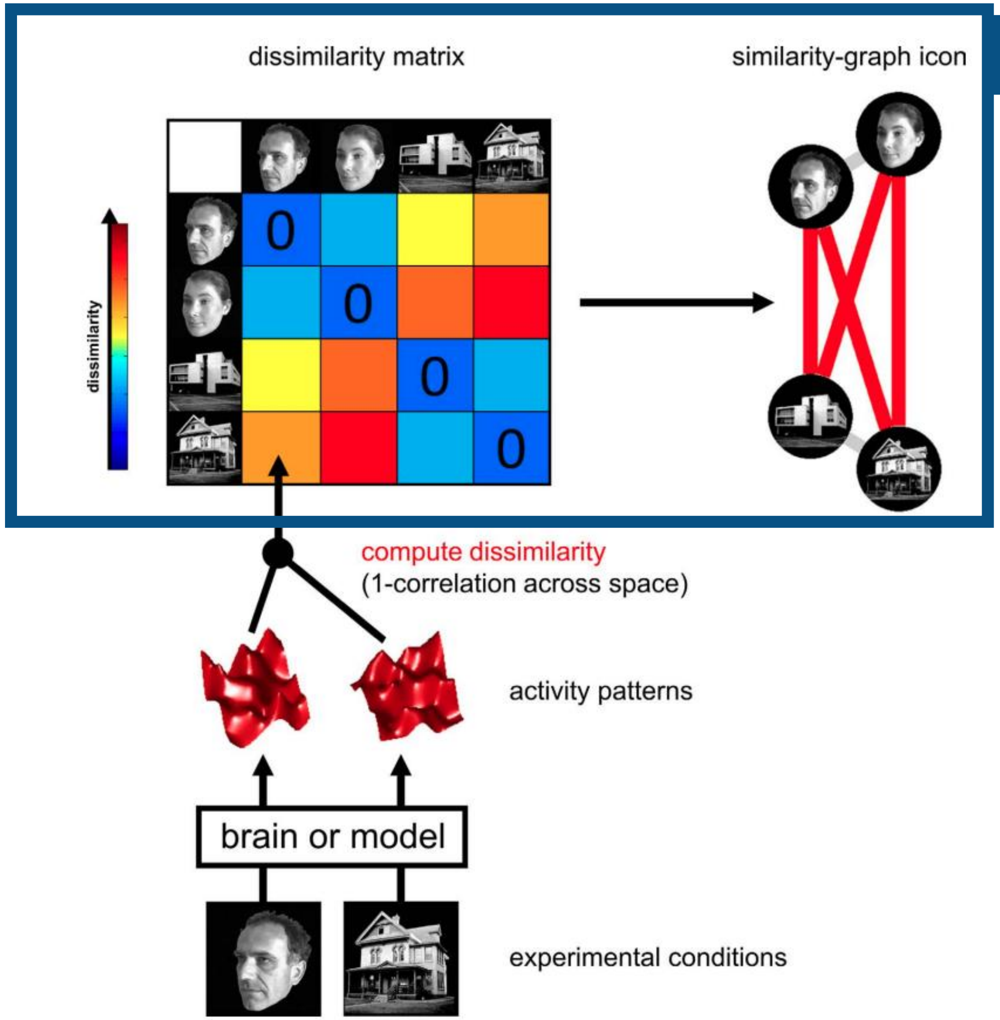


The dissimilarity can be visualised using MDS

I refer to this as the **representational geometry**

Kriegeskorte, Mur, & Bandettini (2008). https://doi.org/10.3389/neuro.06.004.2008

Representational similarity analysis

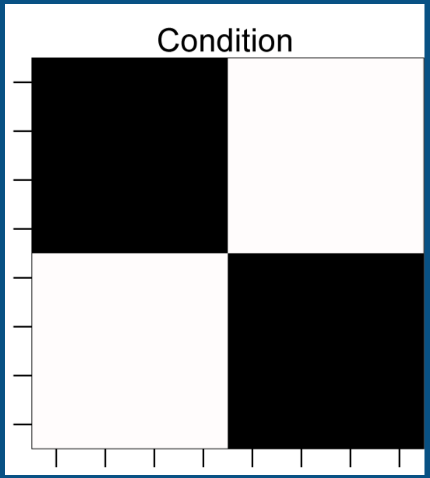


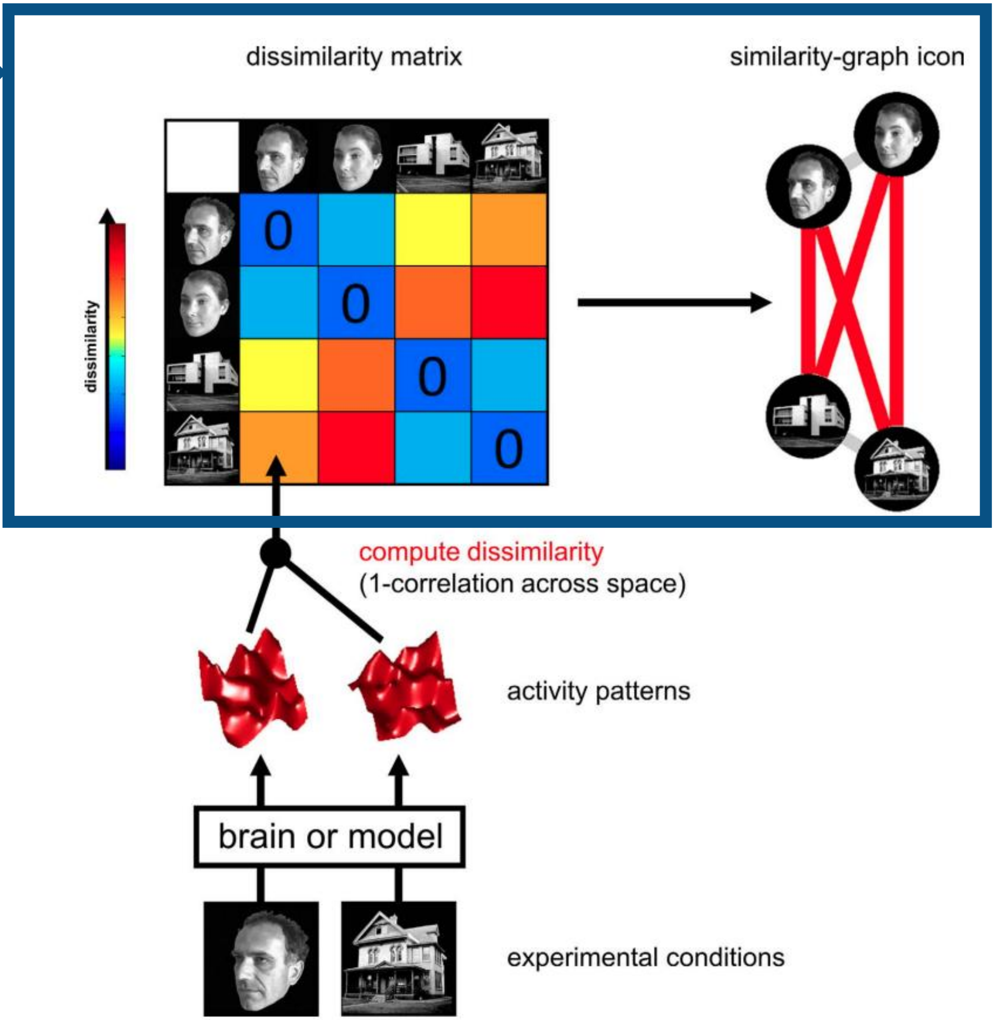
Do these results provide a veridical basis for the representation underlying cognition?

Kriegeskorte, Mur, & Bandettini (2008). https://doi.org/10.3389/neuro.06.004.2008

Representational similarity analysis

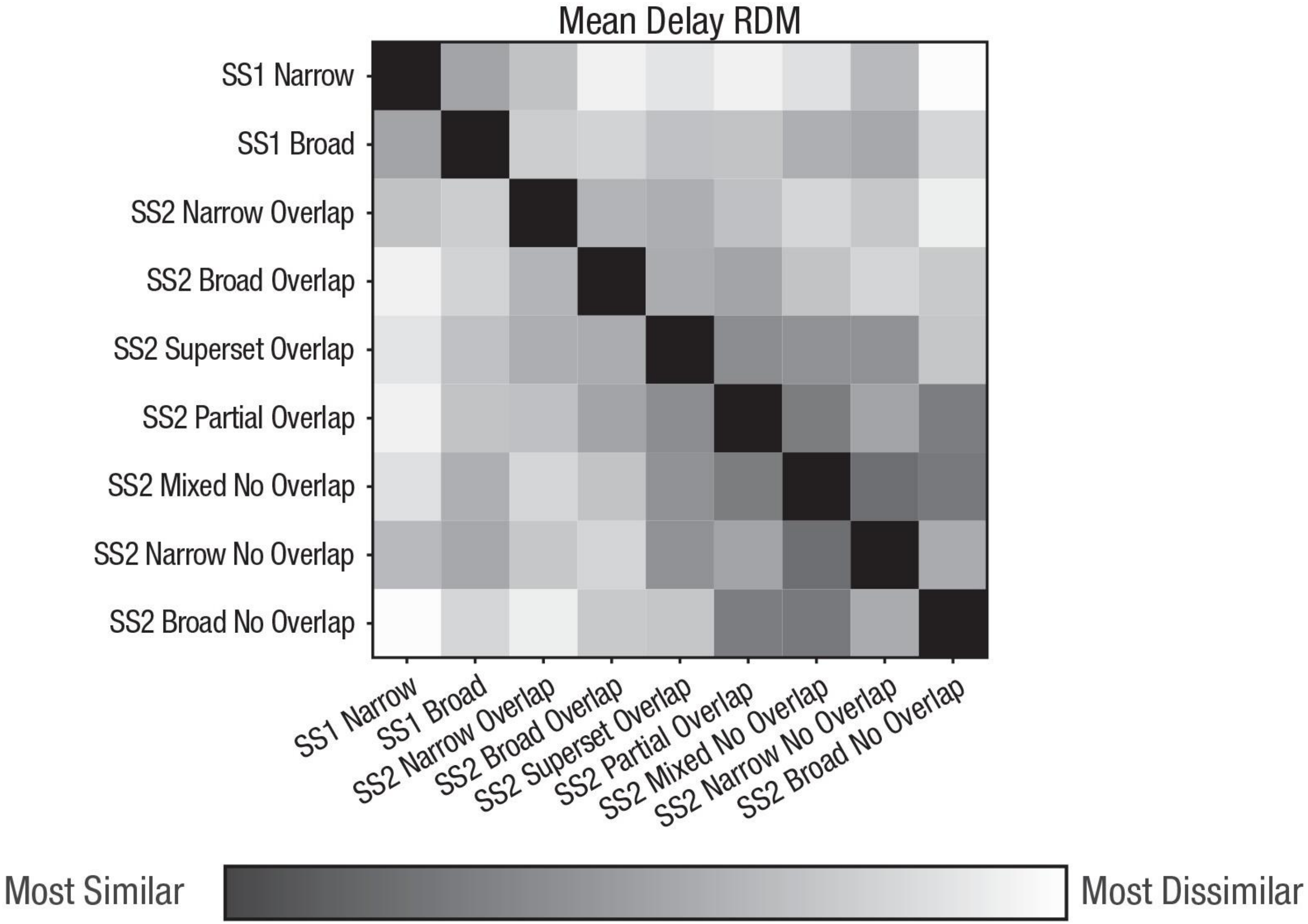
We create models to predict the observed representational (dis)similarity





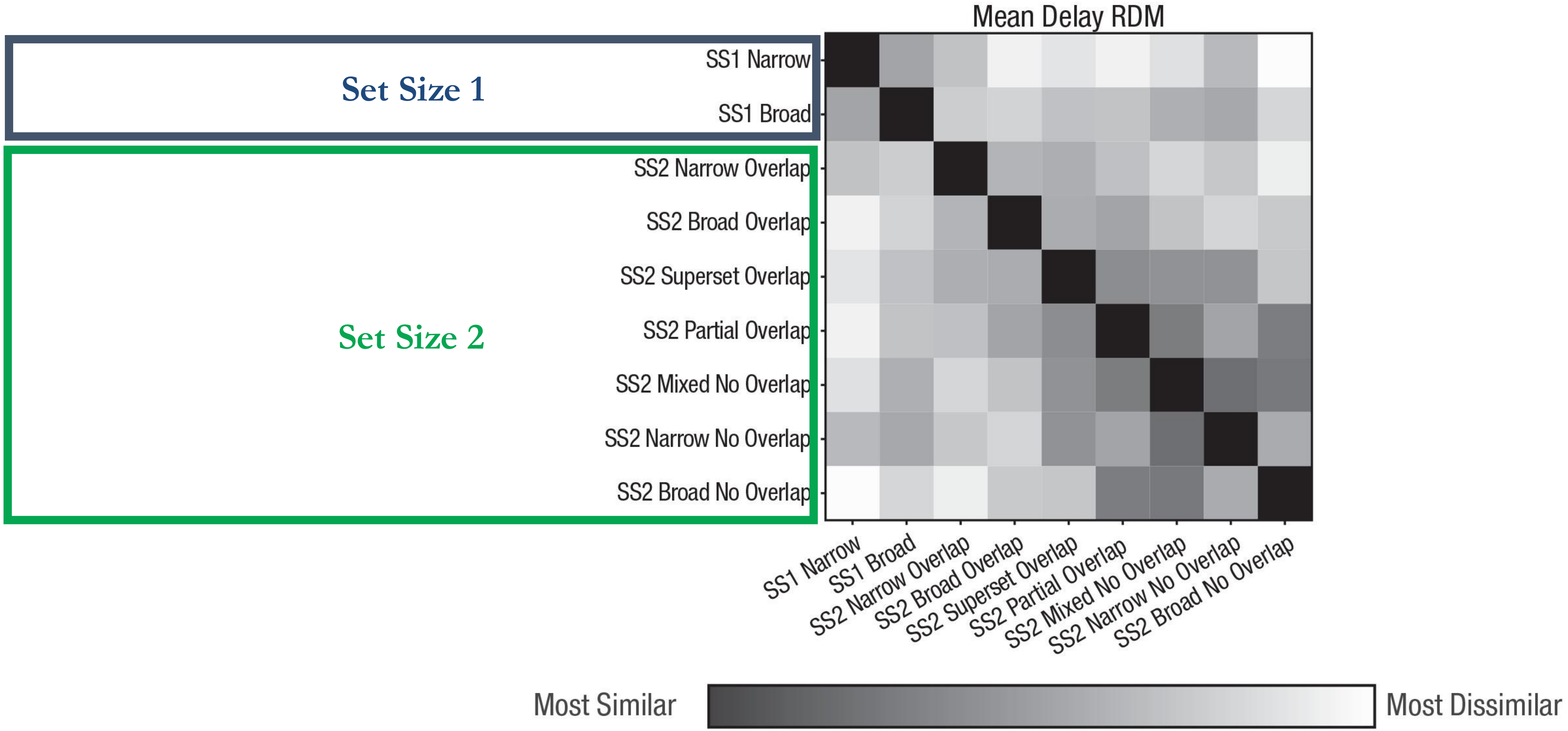
Kriegeskorte, Mur, & Bandettini (2008). https://doi.org/10.3389/neuro.06.004.2008

Decoding load in working memory



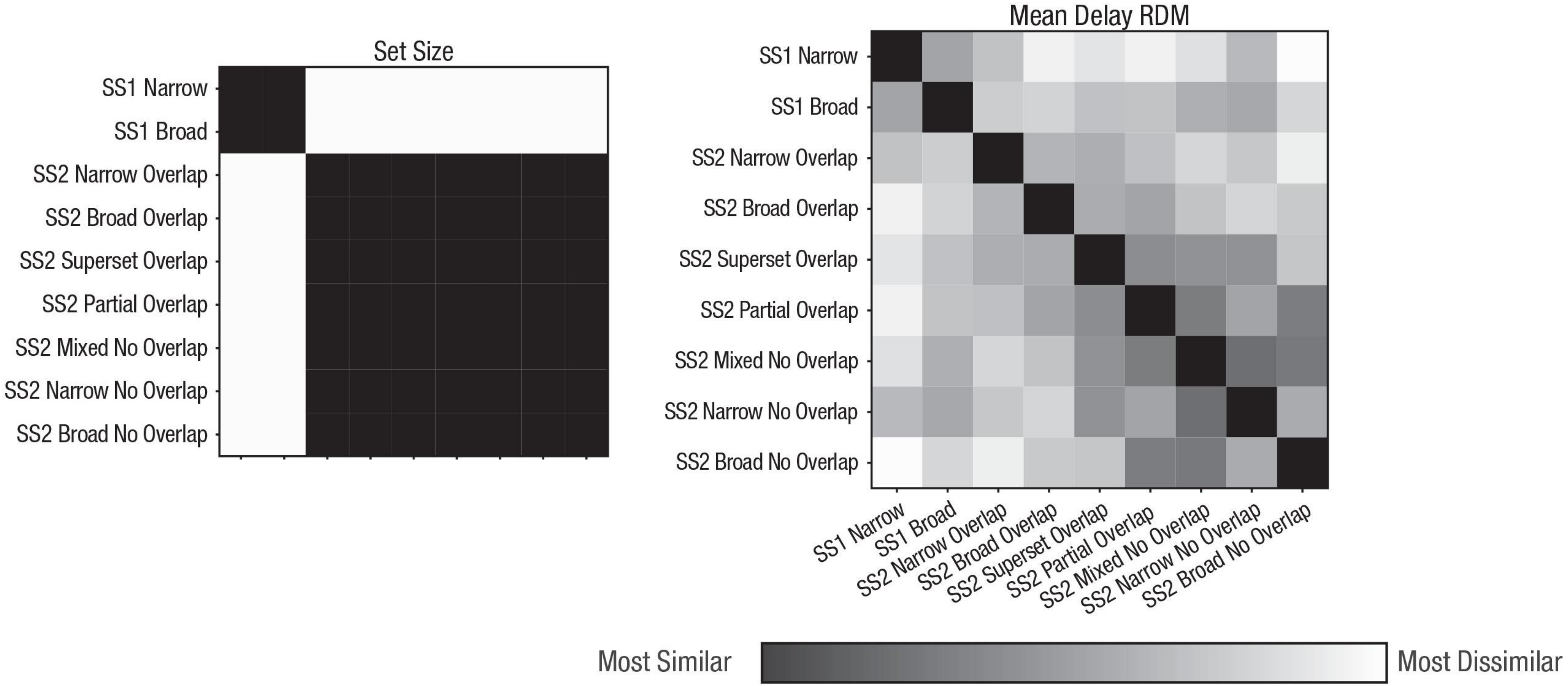
Jones, H. M., Diaz, G. K., Ngiam, W. X. Q., & Awh, E. (2024). Psychological Science, 35(10), 1108-1138.

Decoding load in working memory



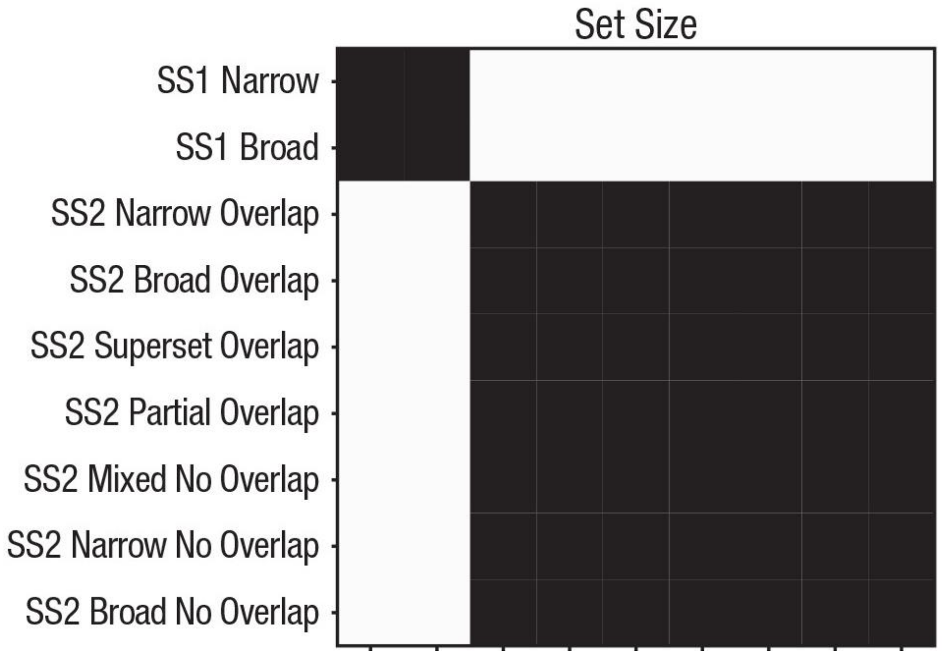
Jones, H. M., Diaz, G. K., Ngiam, W. X. Q., & Awh, E. (2024). Psychological Science, 35(10), 1108-1138.

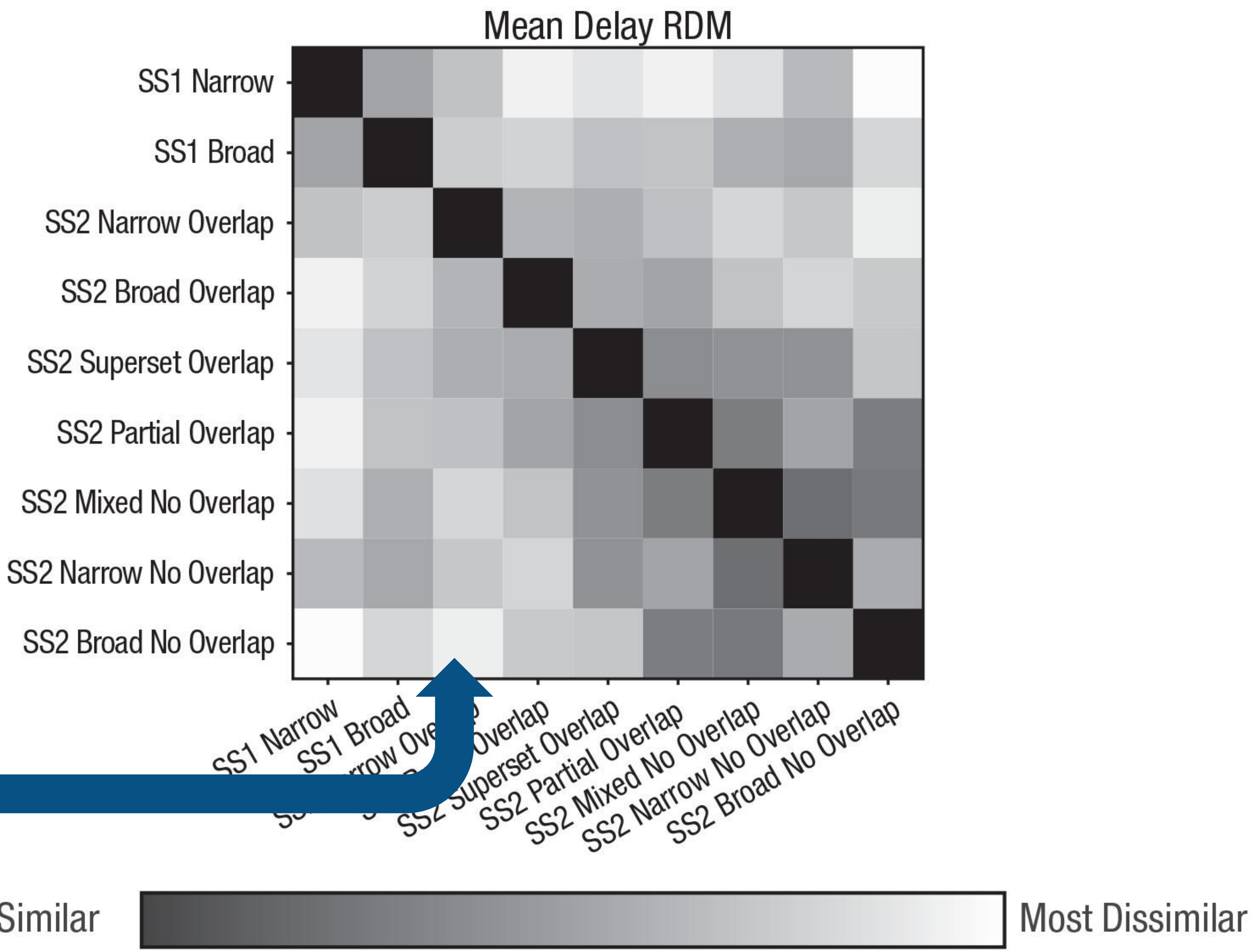
Decoding load in working memory



Jones, H. M., Diaz, G. K., Ngiam, W. X. Q., & Awh, E. (2024). Psychological Science, 35(10), 1108-1138.

Decoding load in working memory

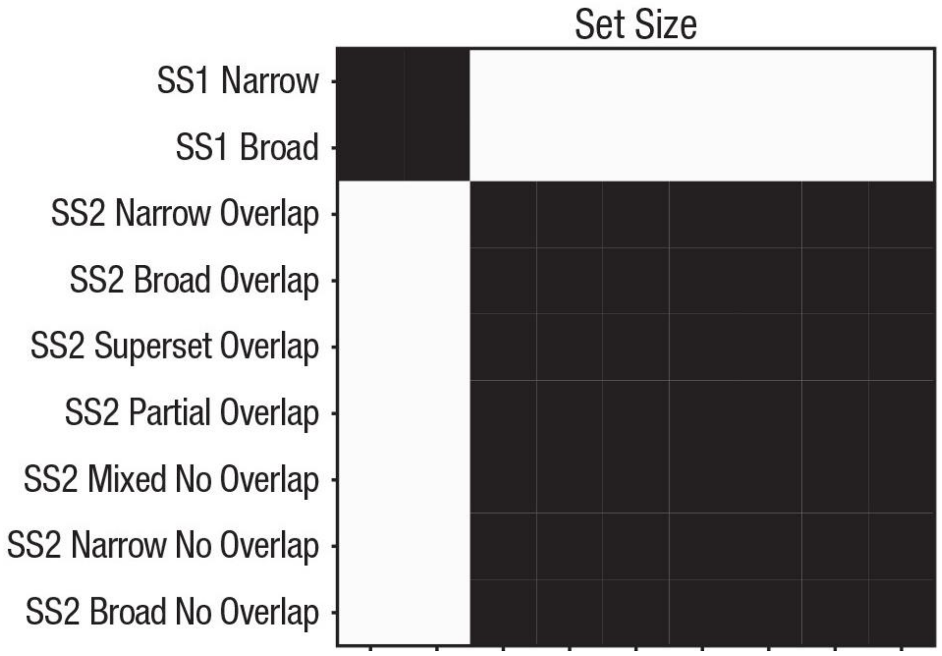




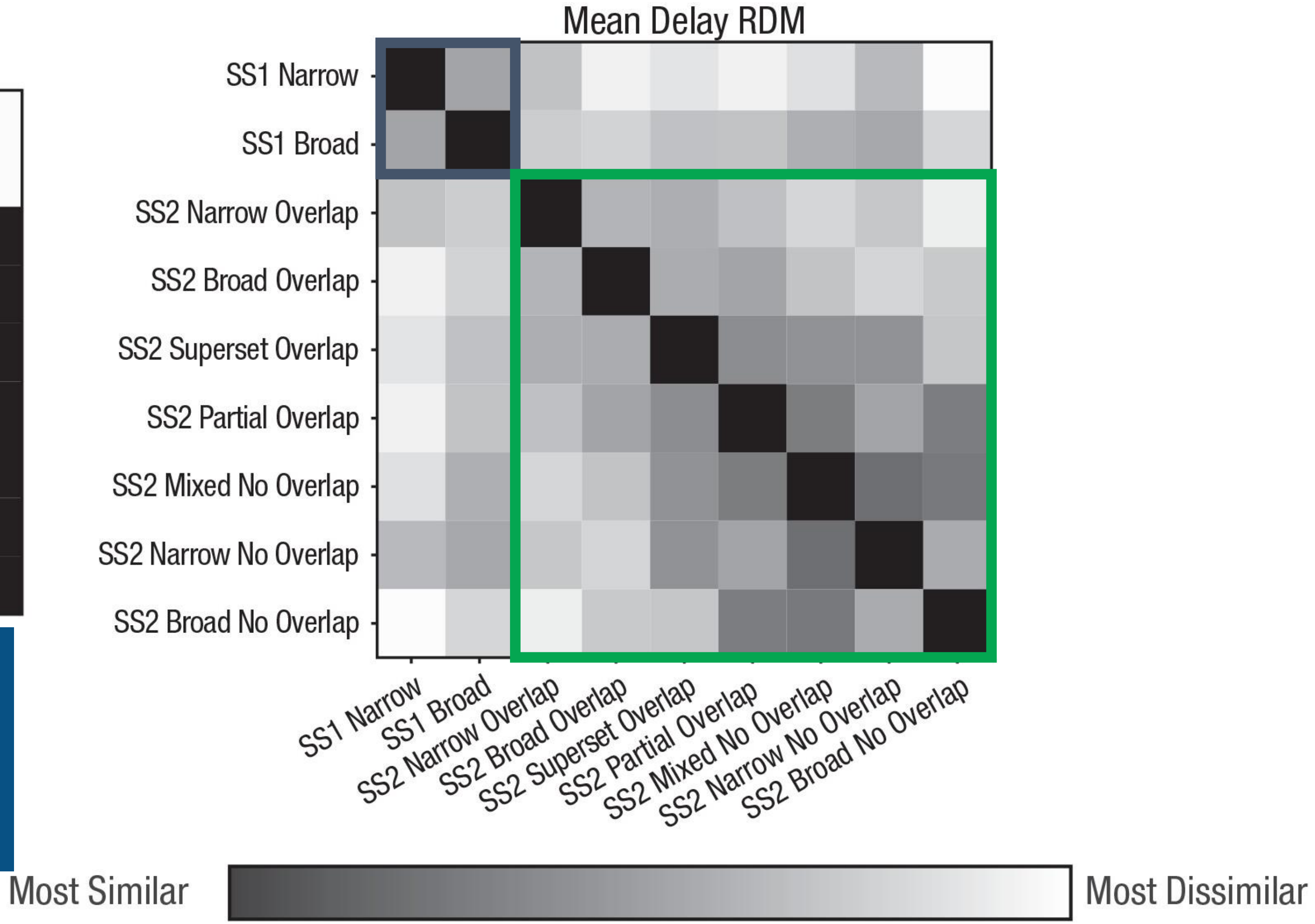
A "model" to predict neural representational (dis)similarity

Jones, H. M., Diaz, G. K., Ngiam, W. X. Q., & Awh, E. (2024). Psychological Science, 35(10), 1108-1138.

Decoding load in working memory



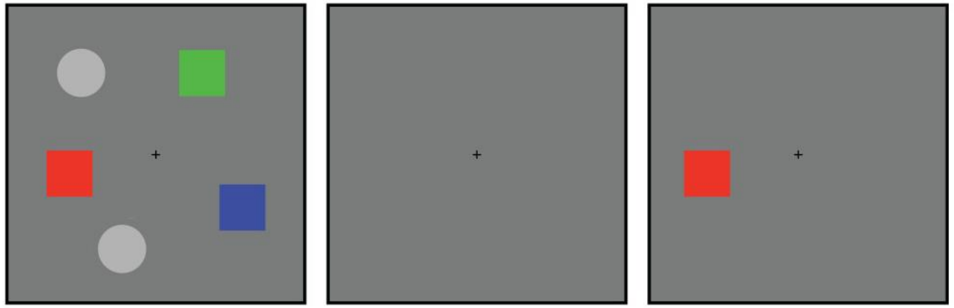
A "model" to predict neural representational (dis)similarity



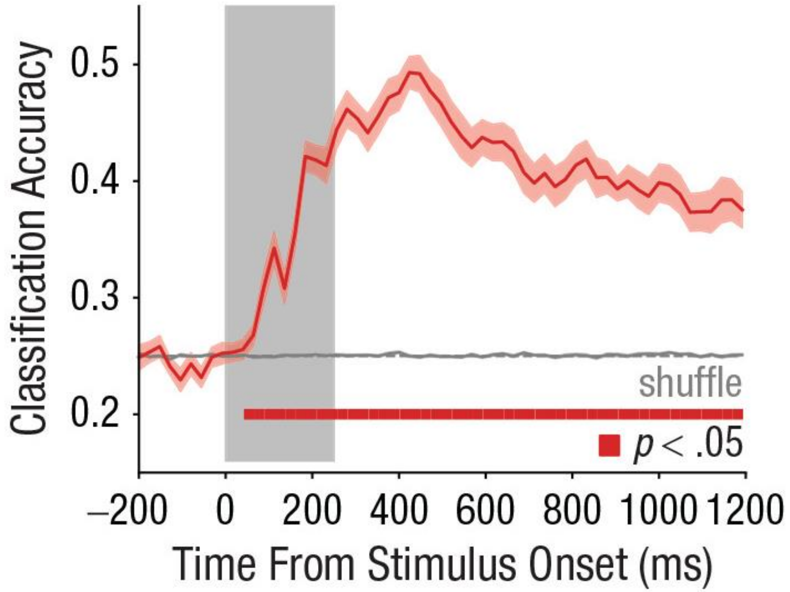
Jones, H. M., Diaz, G. K., Ngiam, W. X. Q., & Awh, E. (2024). Psychological Science, 35(10), 1108-1138.

Decoding load in working memory

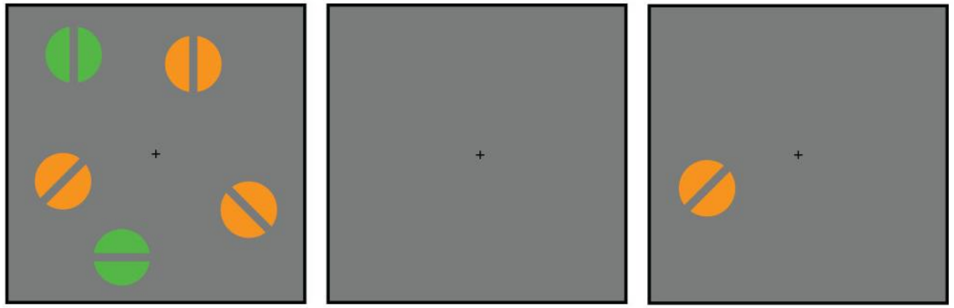
Experiment 1: Color



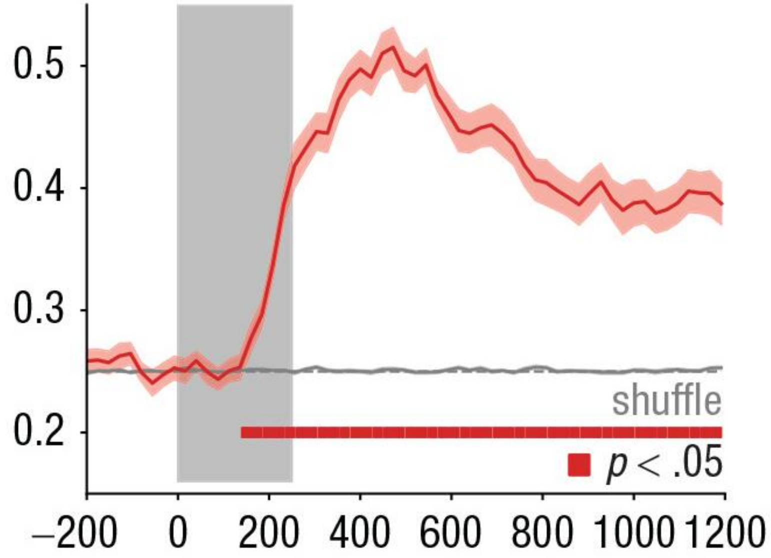
Train and test

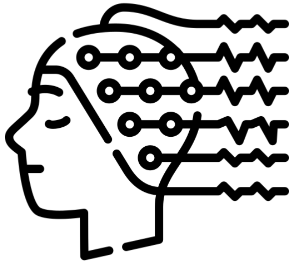


Experiment 2: Orientation



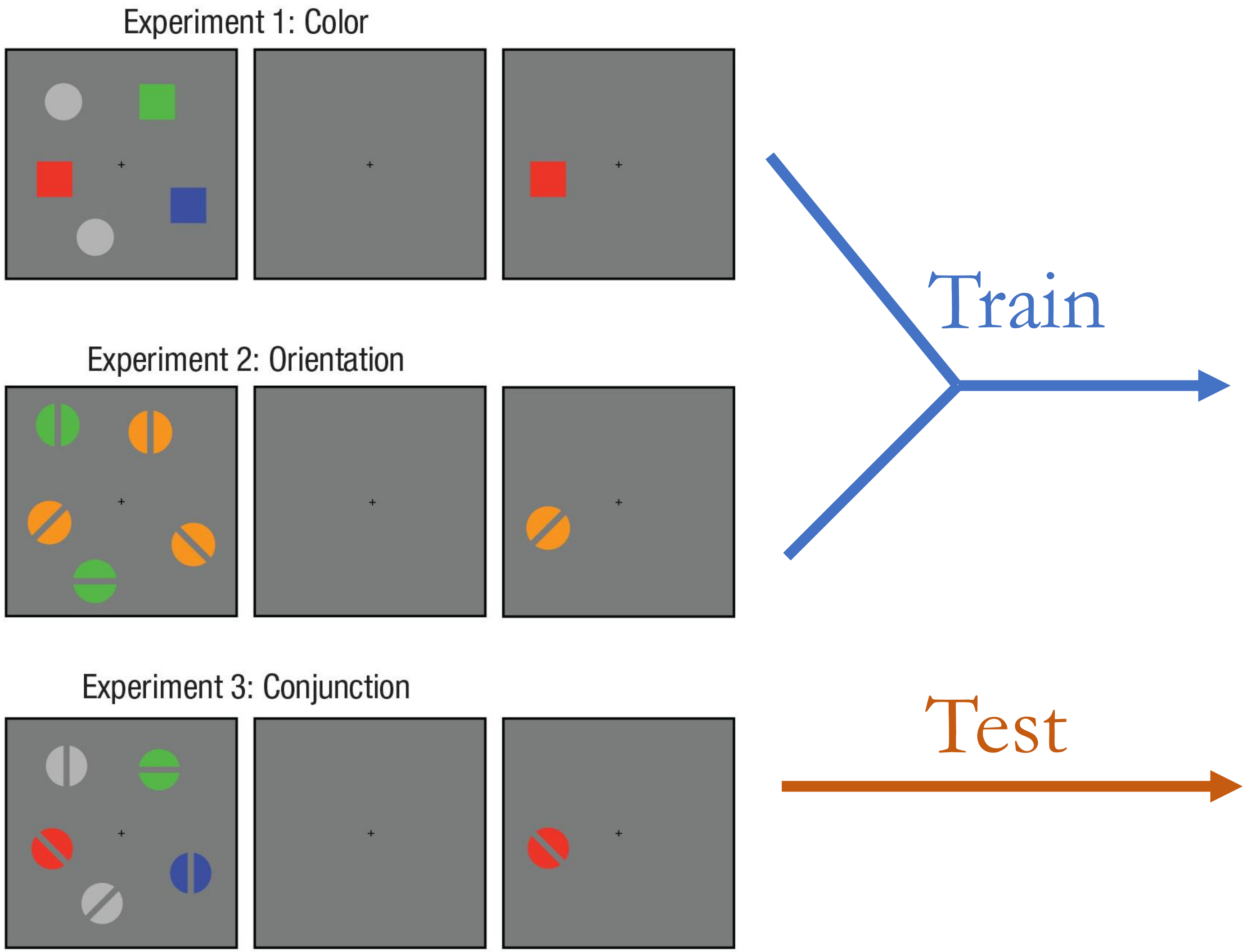
Train and test

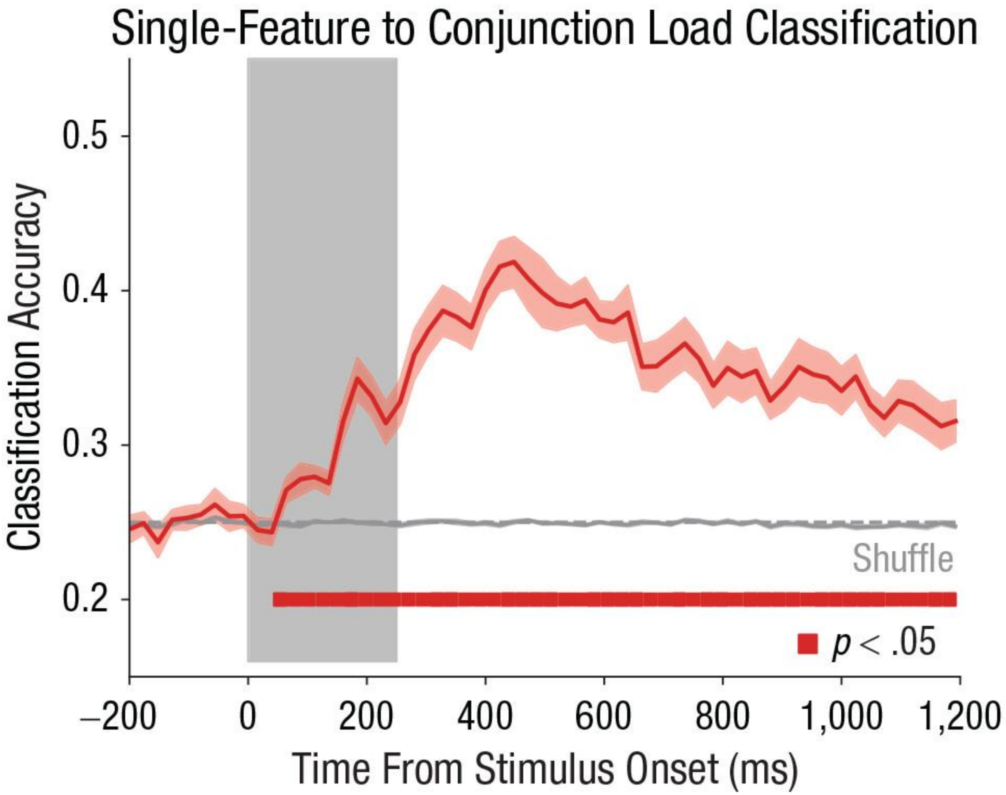




Thyer et al. (2022) Psychological Science

Decoding load in working memory

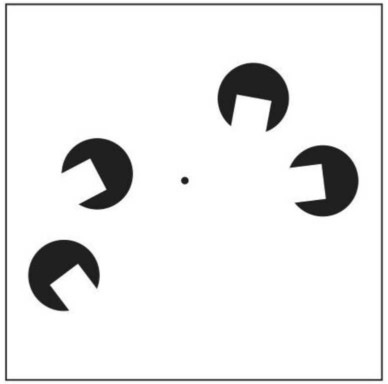




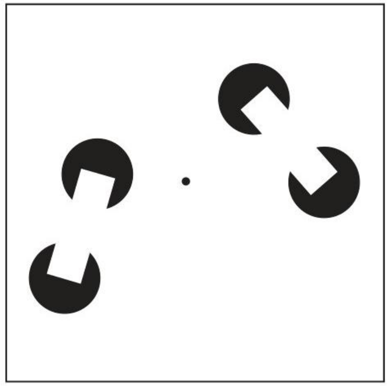
Thyer et al. (2022) Psychological Science

Decoding load in working memory

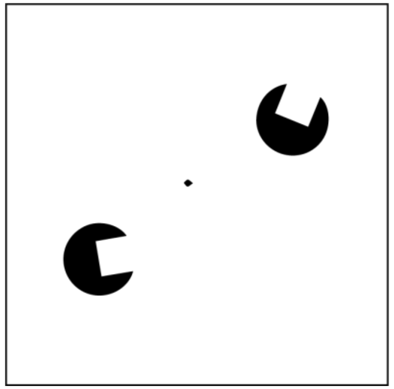
4 Ungrouped

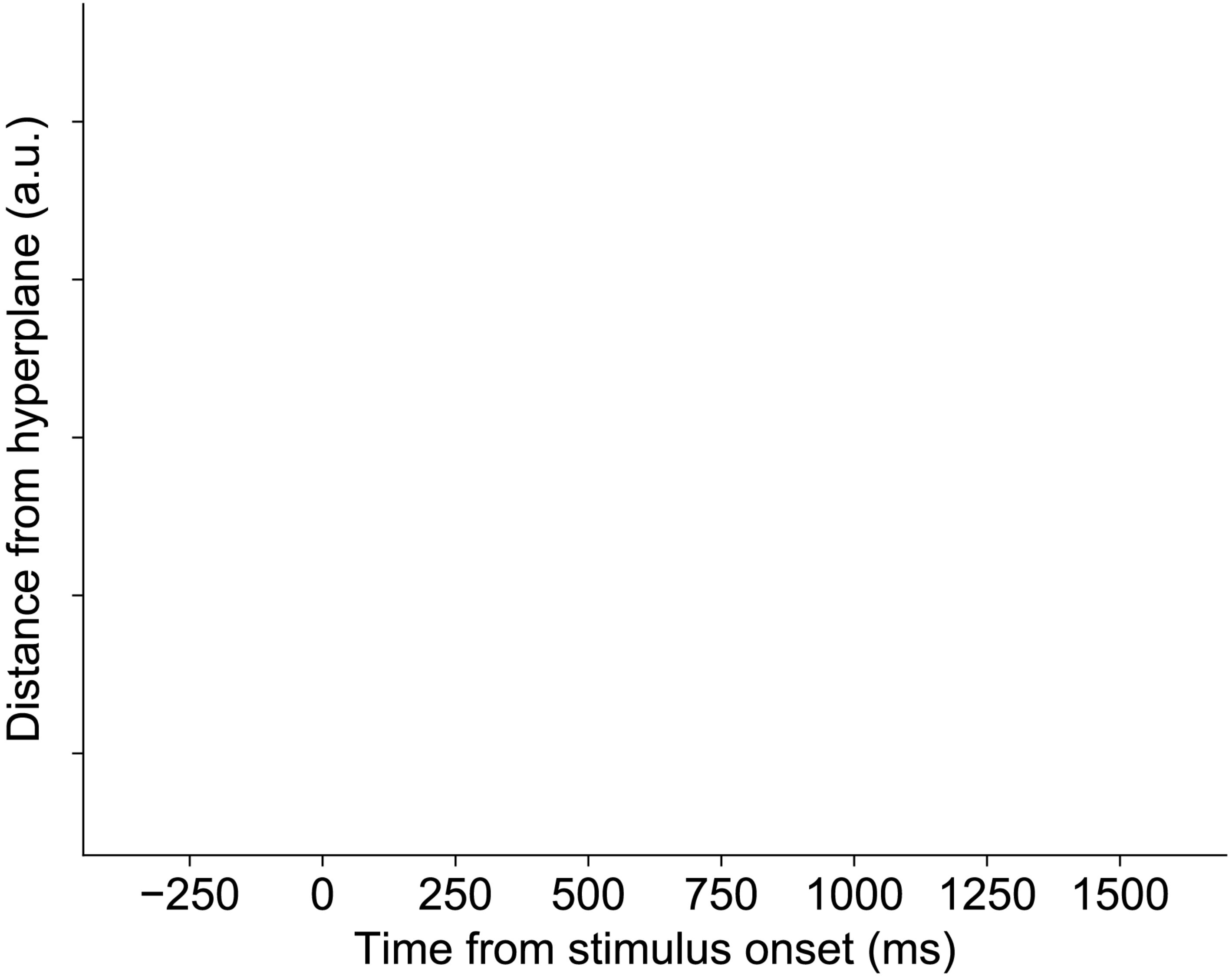


4 Grouped



2 Ungrouped

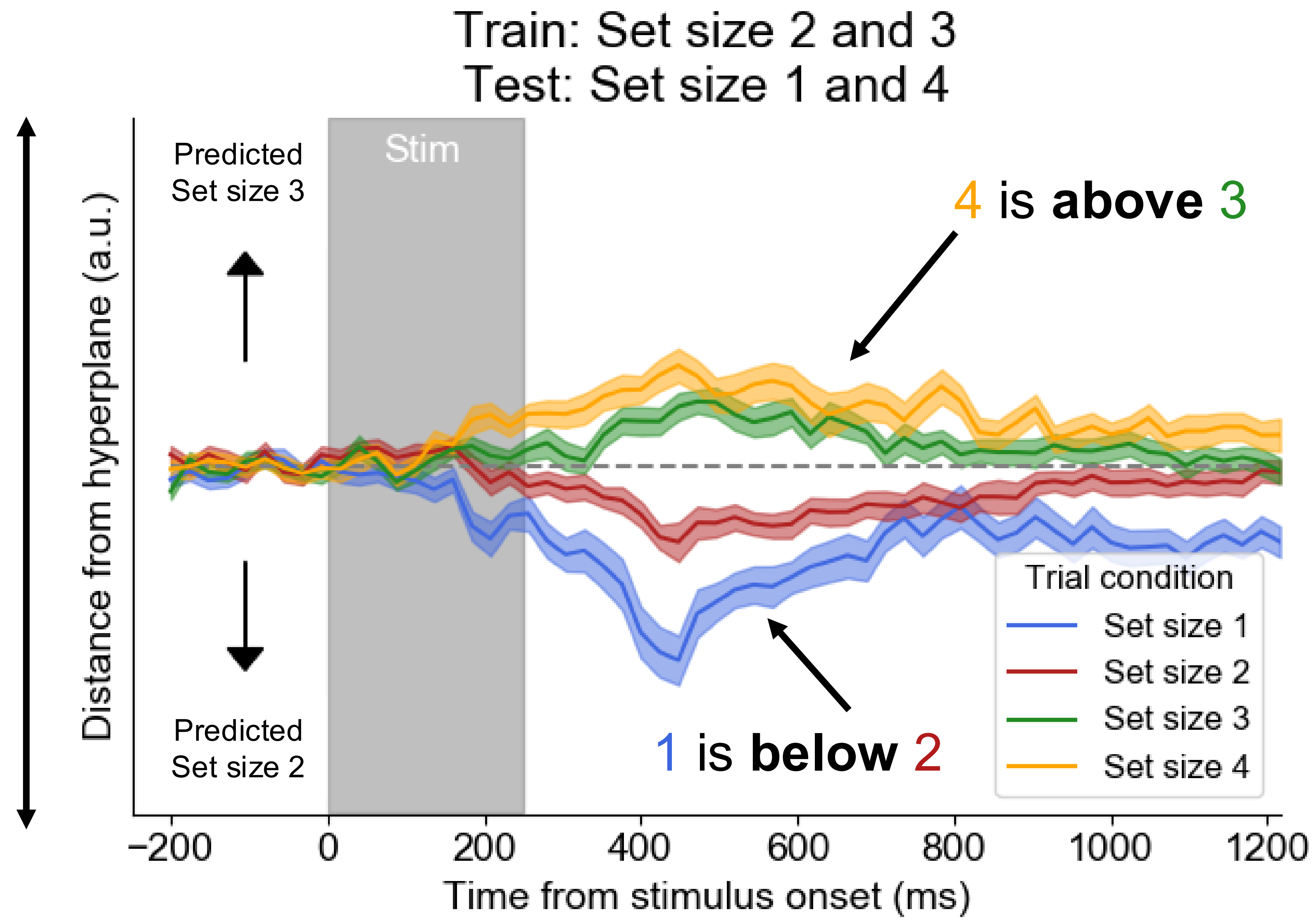




Thyer et al. (2022) Psychological Science

Decoding load in working memory

Working memory load signal



Decoding load in working memory

We were interested in how associative learning influences working memory operations.

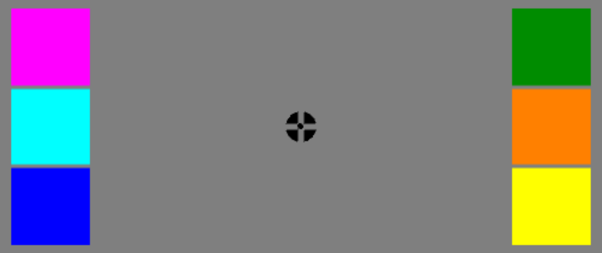
One proposed operation is that associative learning leads to "chunking" processes – representations where separate items are bound into a "chunk".

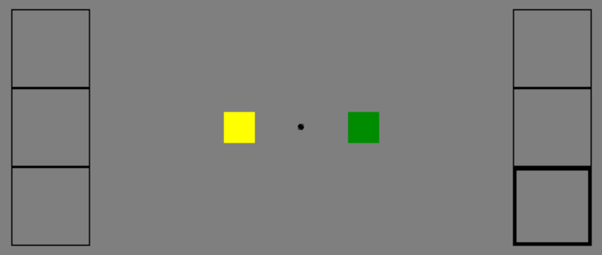




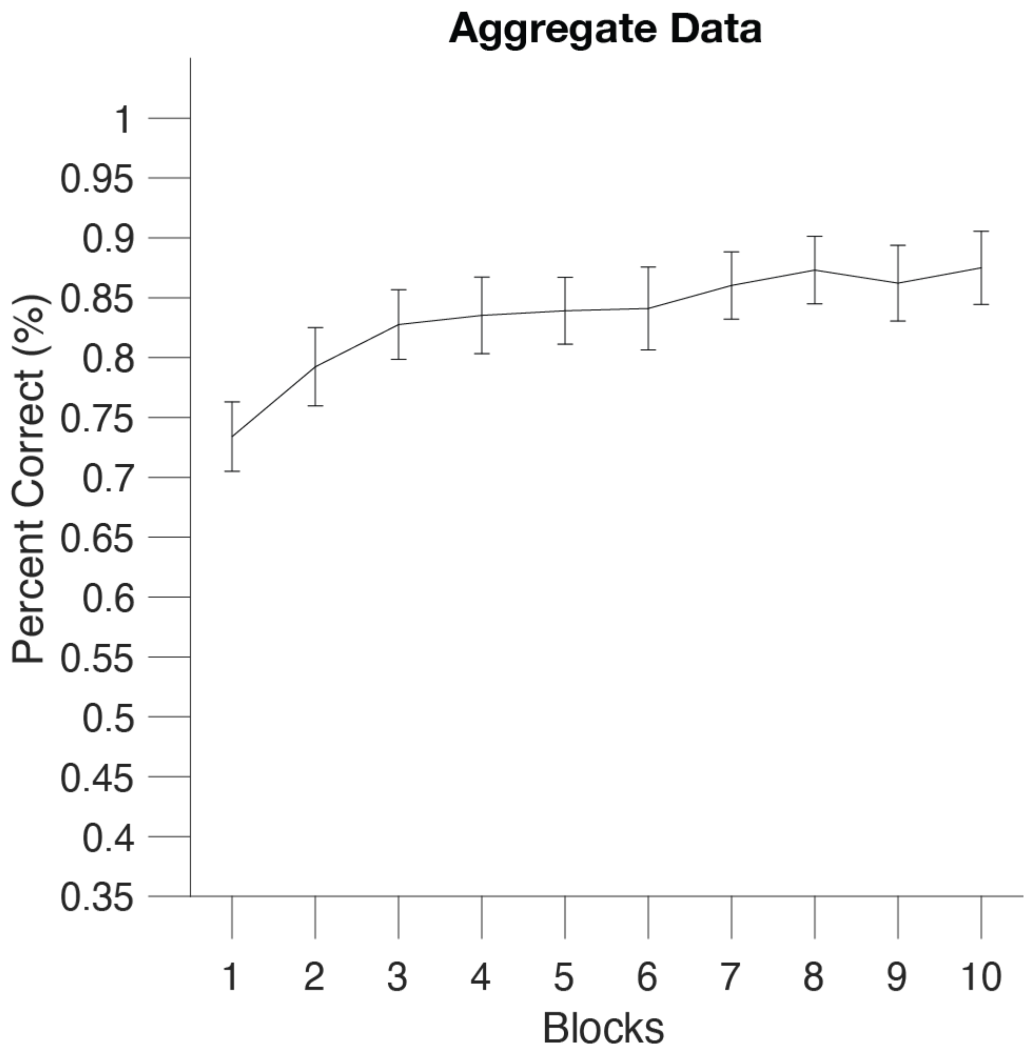
Training

Trained subjects to learn three color triplets

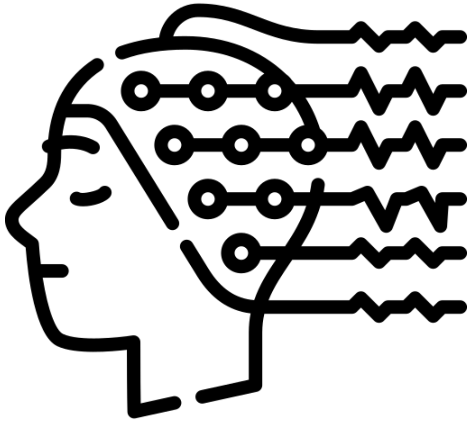


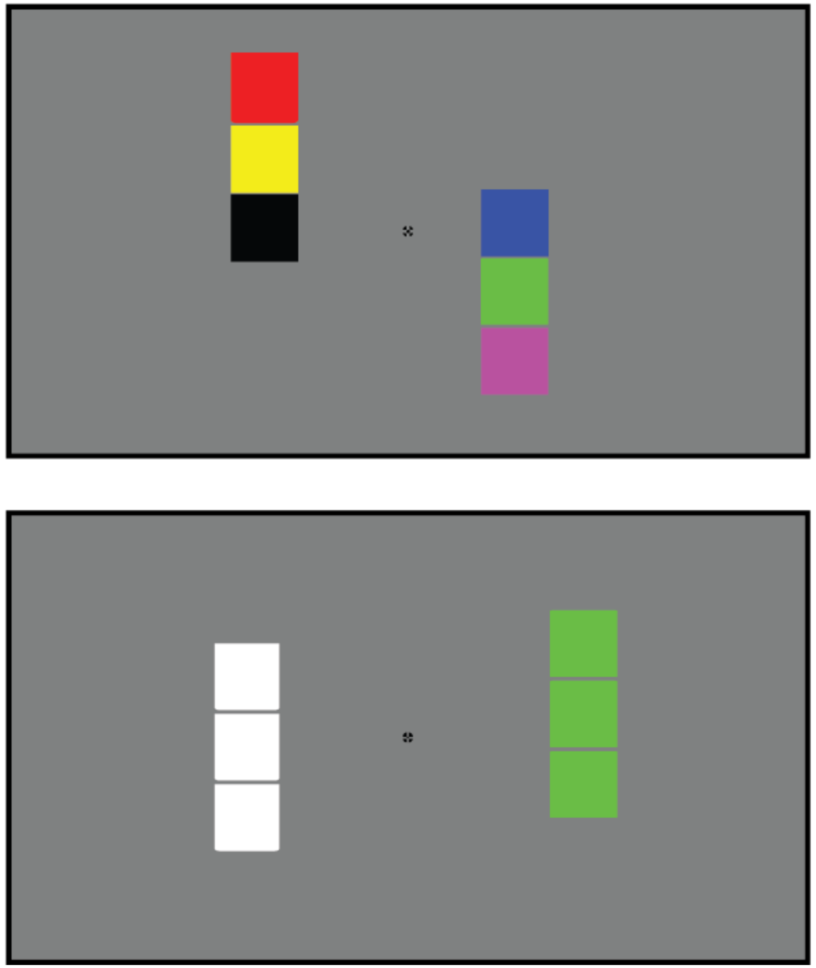


Training Results



EEG



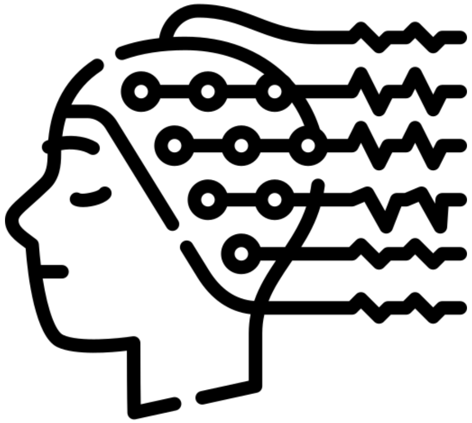


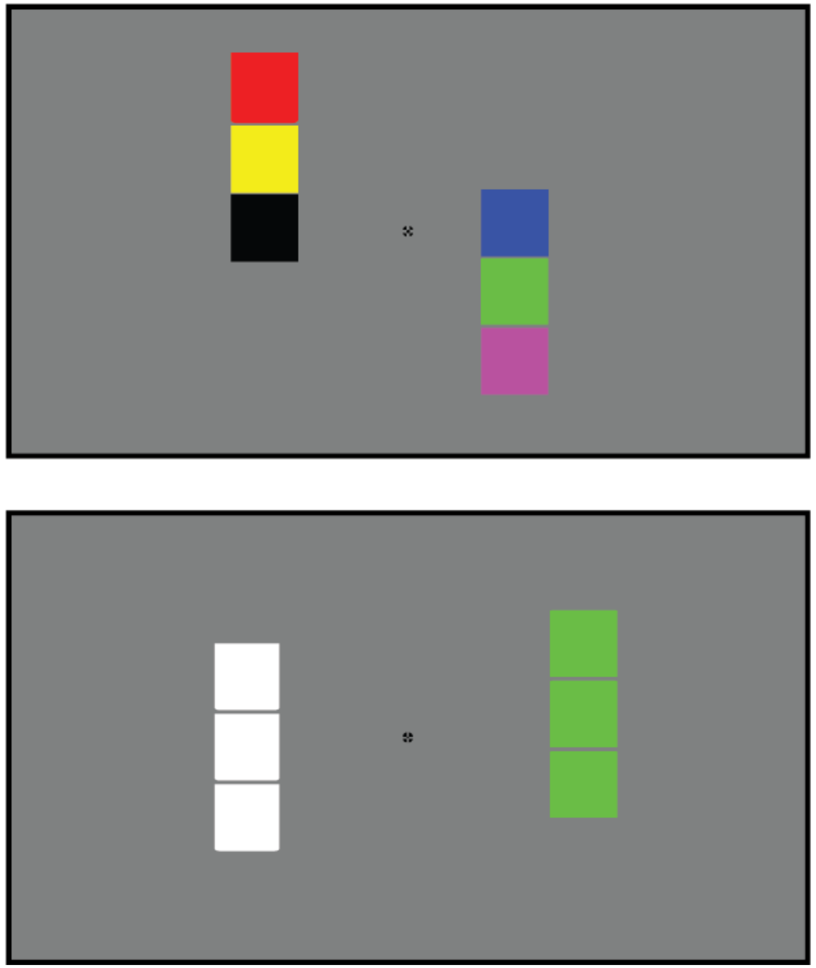
Perceptually
equivalent

Six random
Six chunked

Two random

EEG





Perceptually equivalent

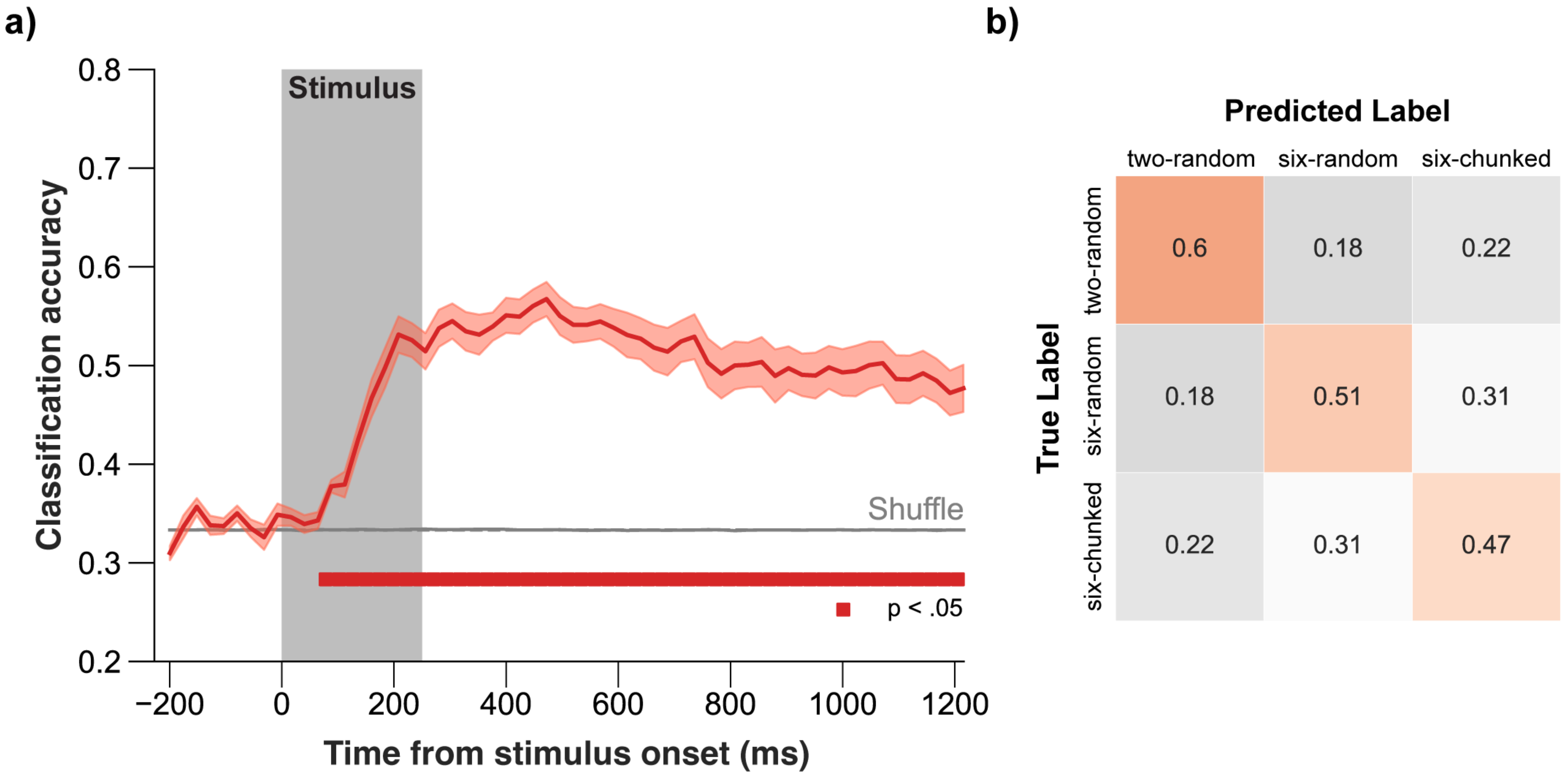
Six random

Six chunked

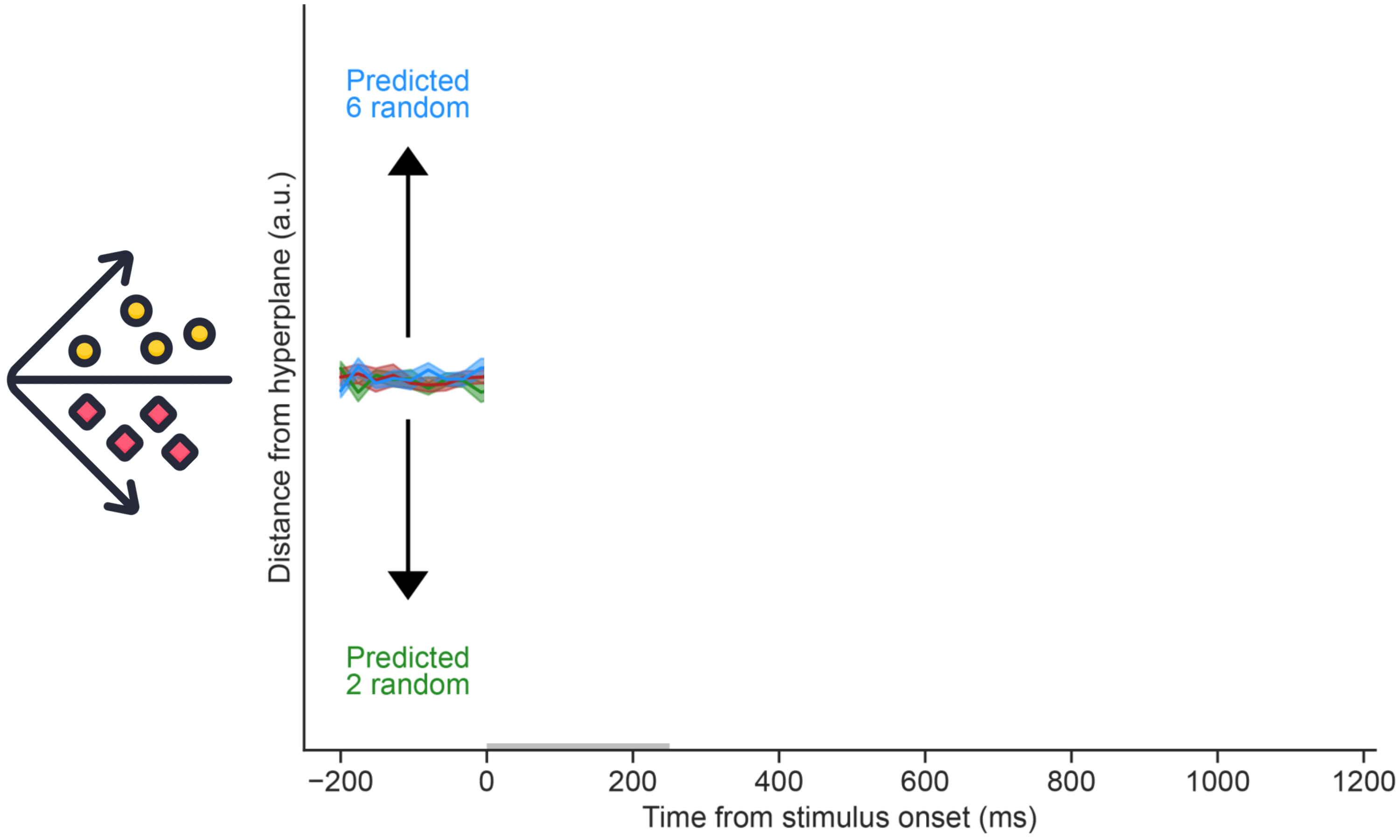
Two random

Expectation: "chunking" results in a reduction of item-based load that should be reflected in neural representations

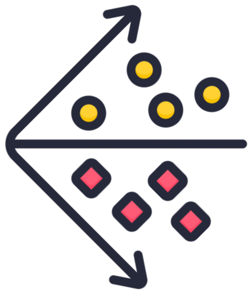
Multivariate classification

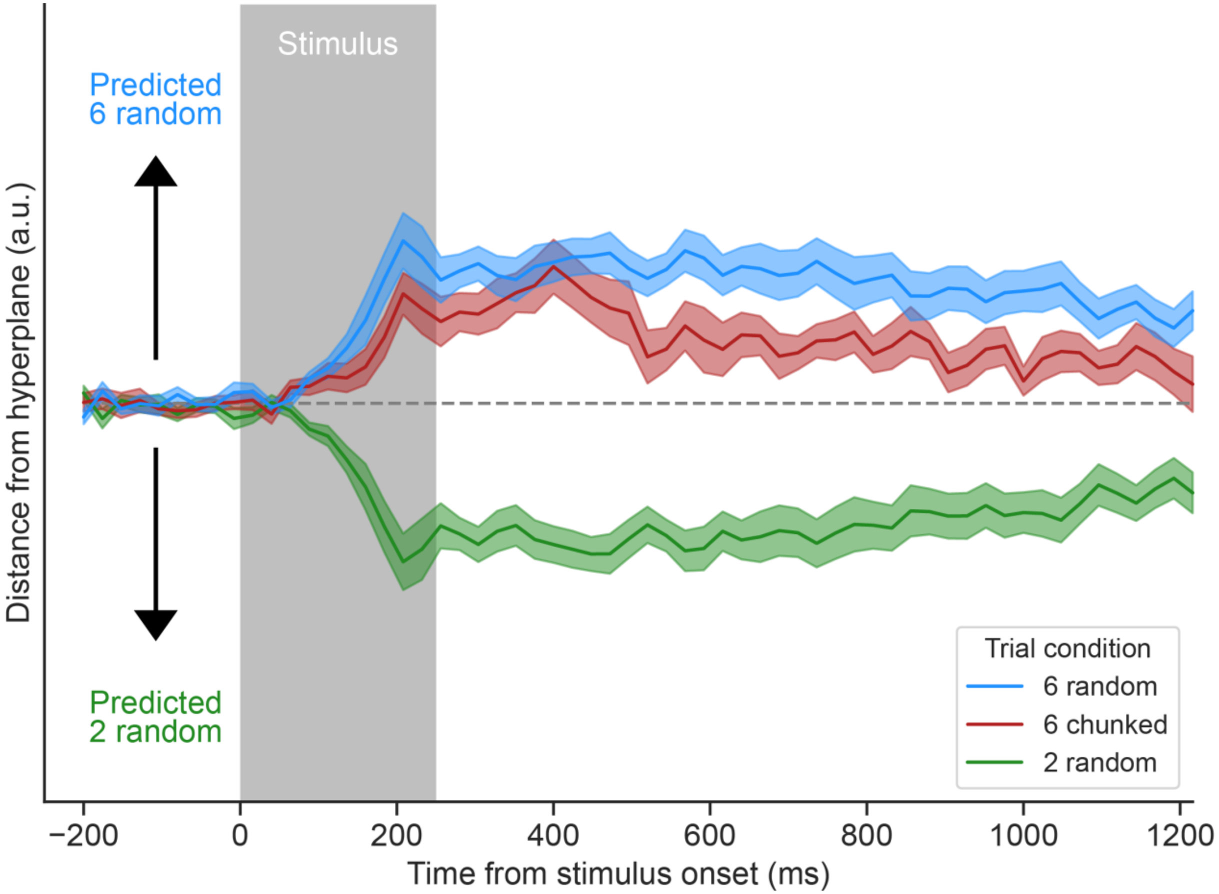


Train 6 random versus 2 random, test 6 chunked

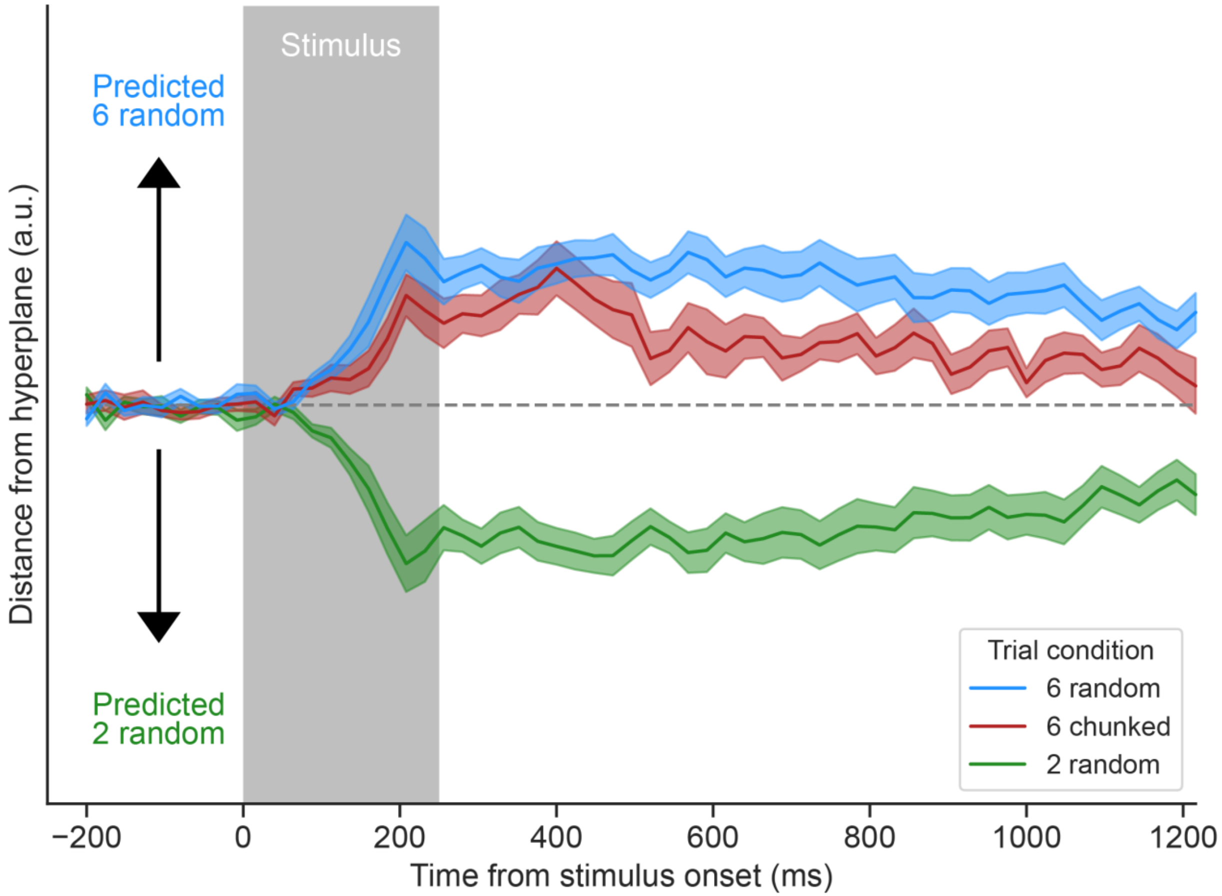


Train 6 random versus 2 random, test 6 chunked

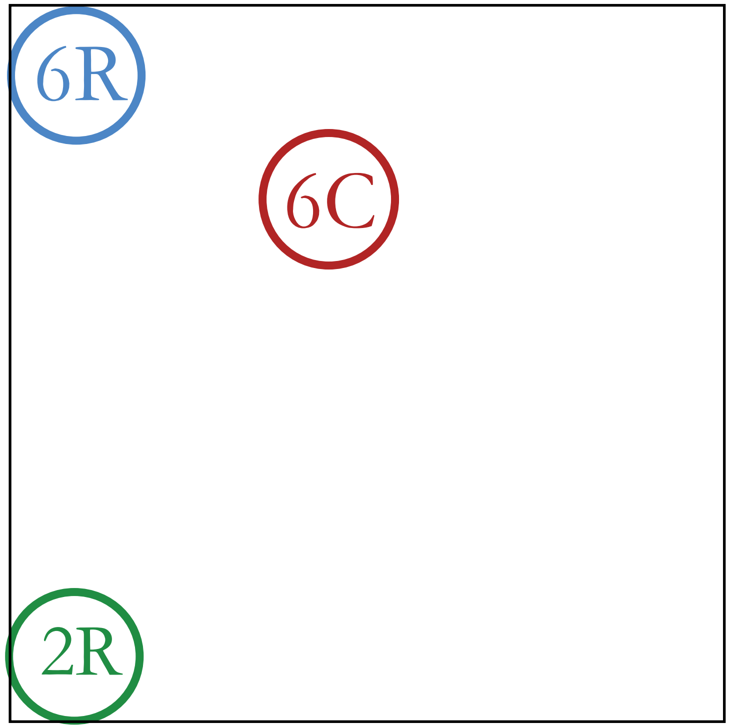




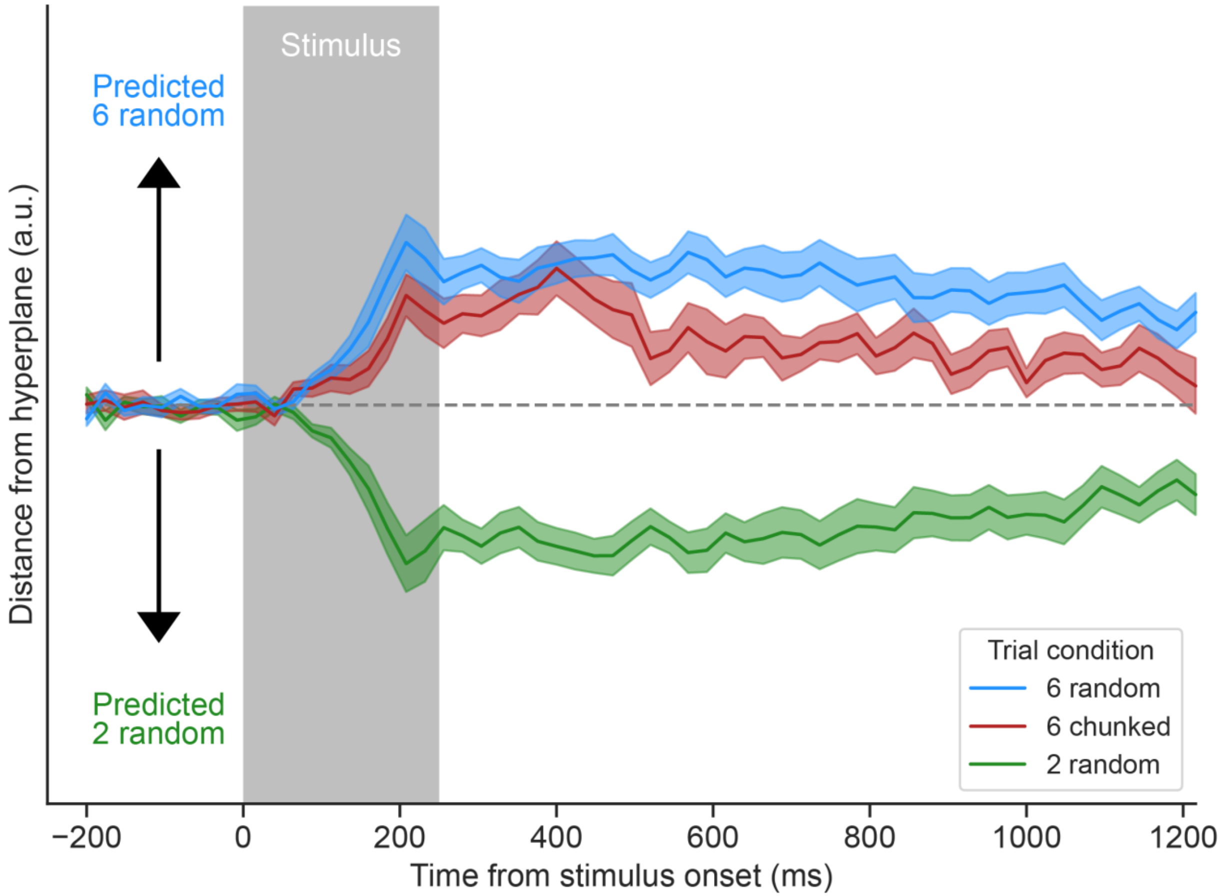
Train 6 random versus 2 random, test 6 chunked

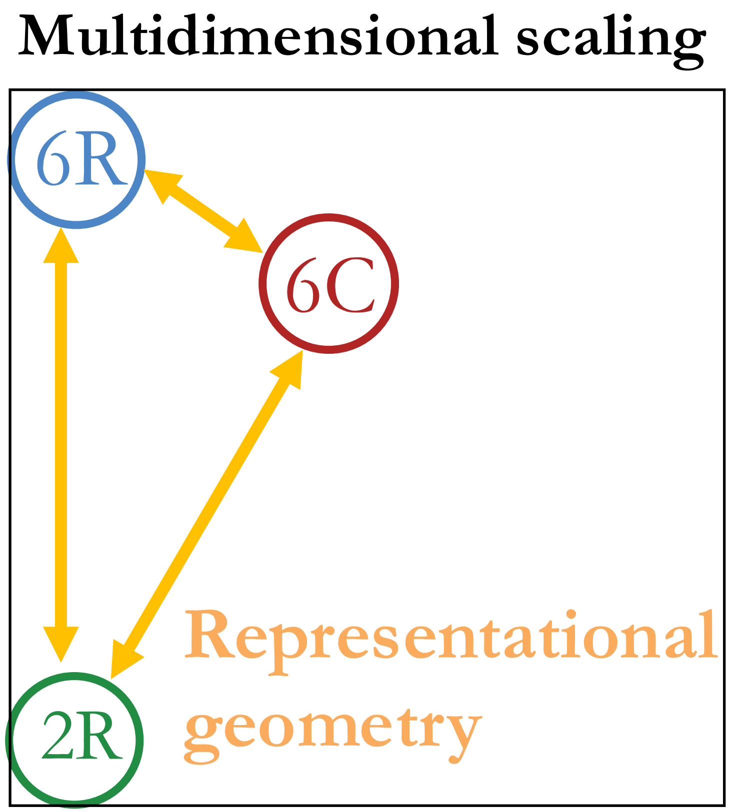


Multidimensional scaling

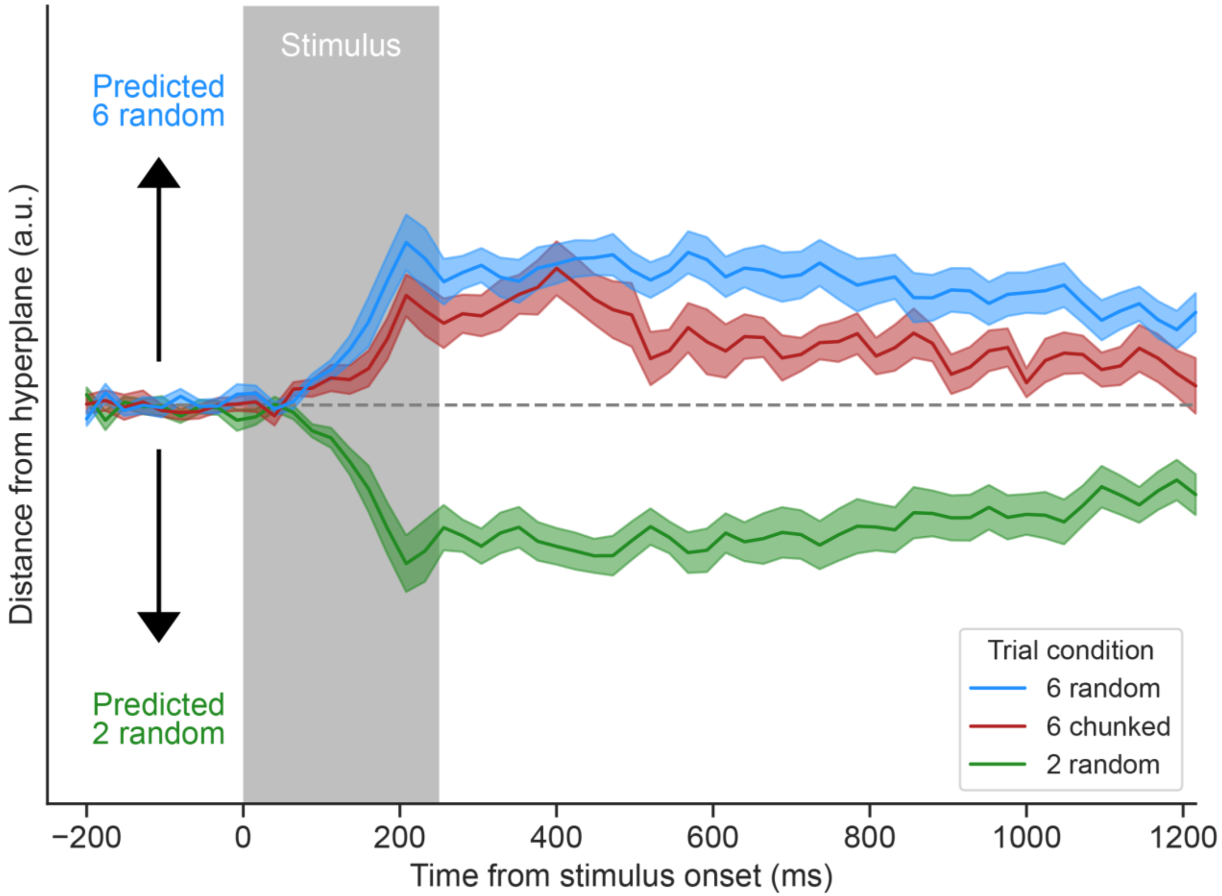


Train 6 random versus 2 random, test 6 chunked

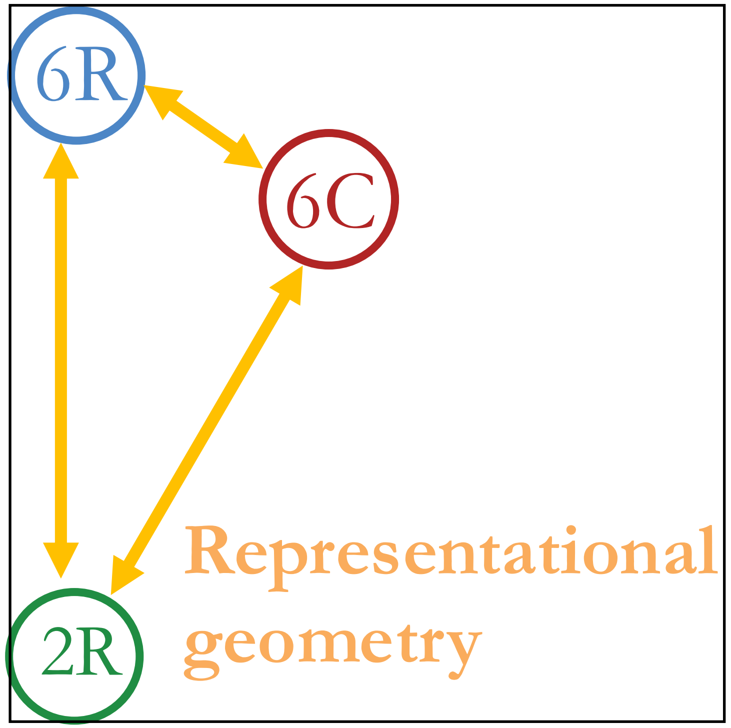




Train 6 random versus 2 random, test 6 chunked

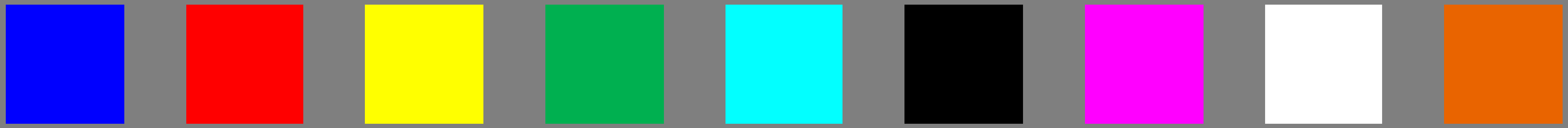


Multidimensional scaling



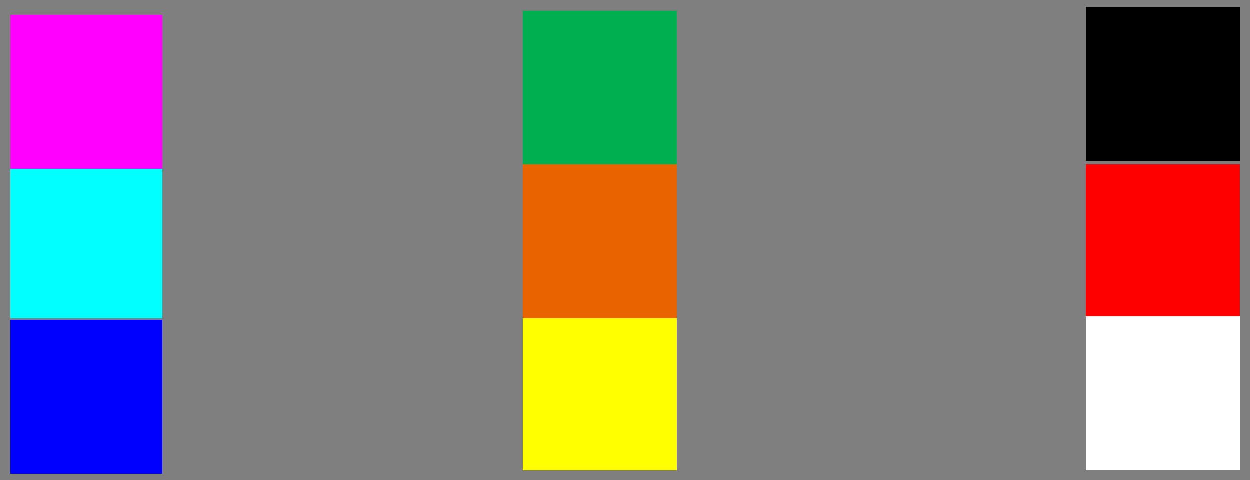
What can we infer from these results?
What cognitive model would predict or could explain this pattern of representational similarity?

Awareness Test

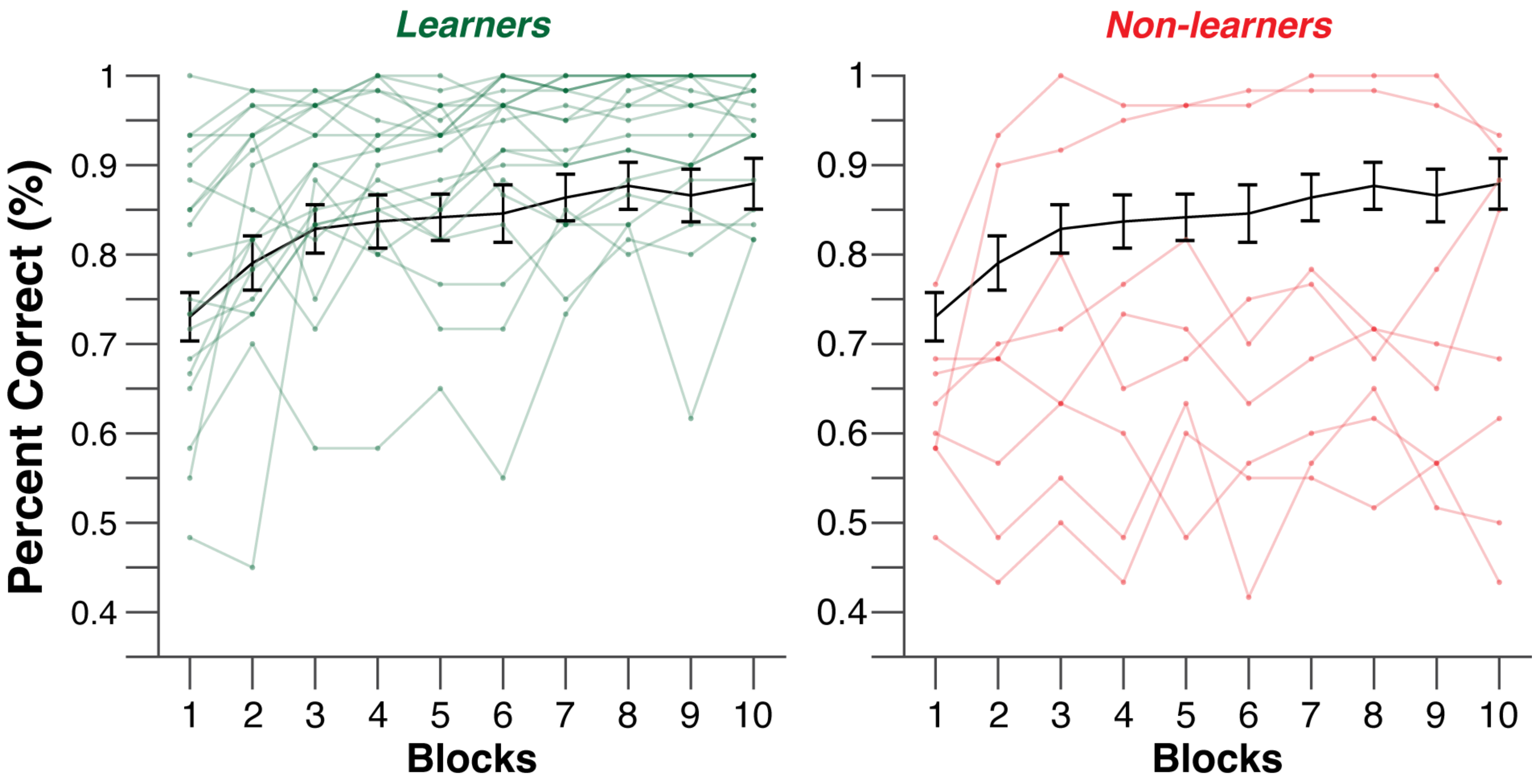


Awareness Test

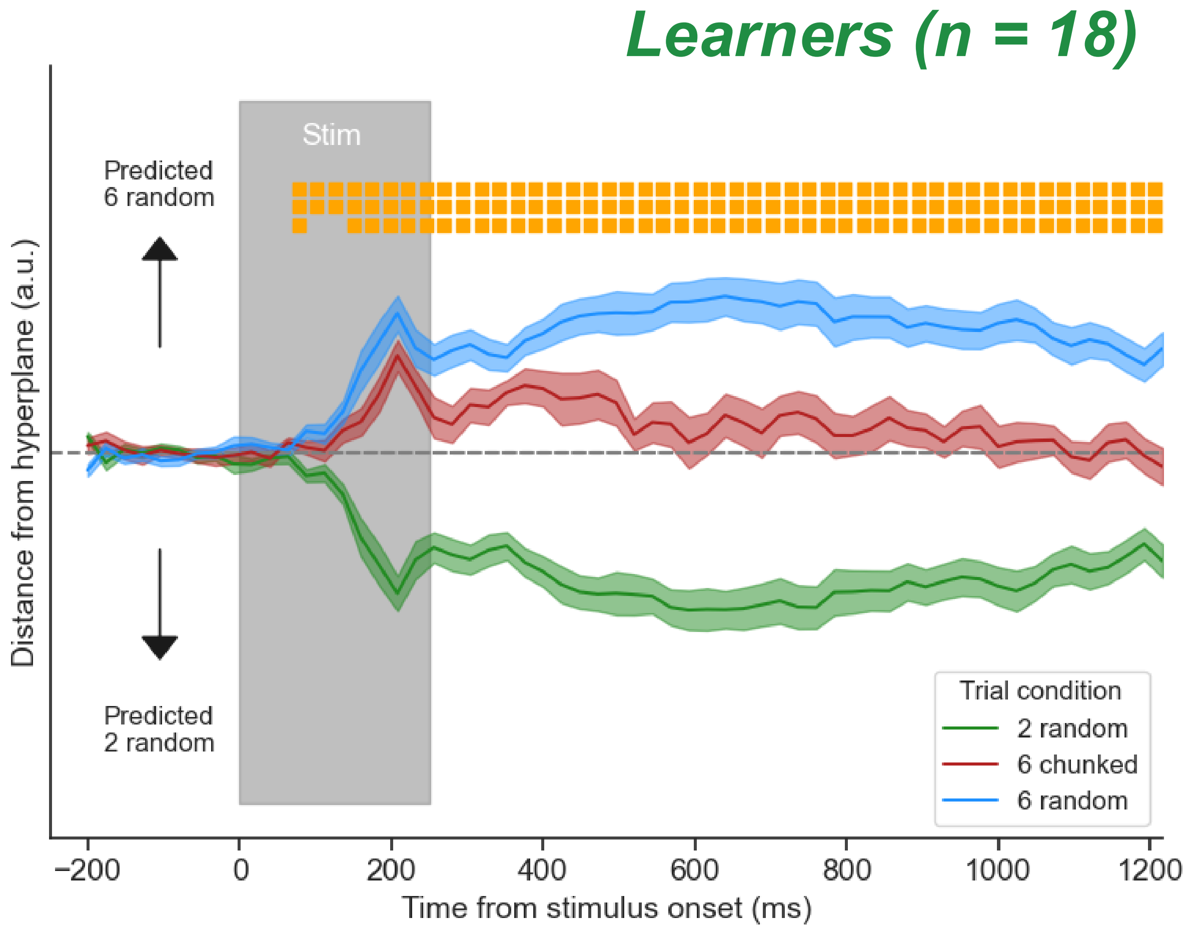
Only subjects that recreated **all** triplets were considered “learners”

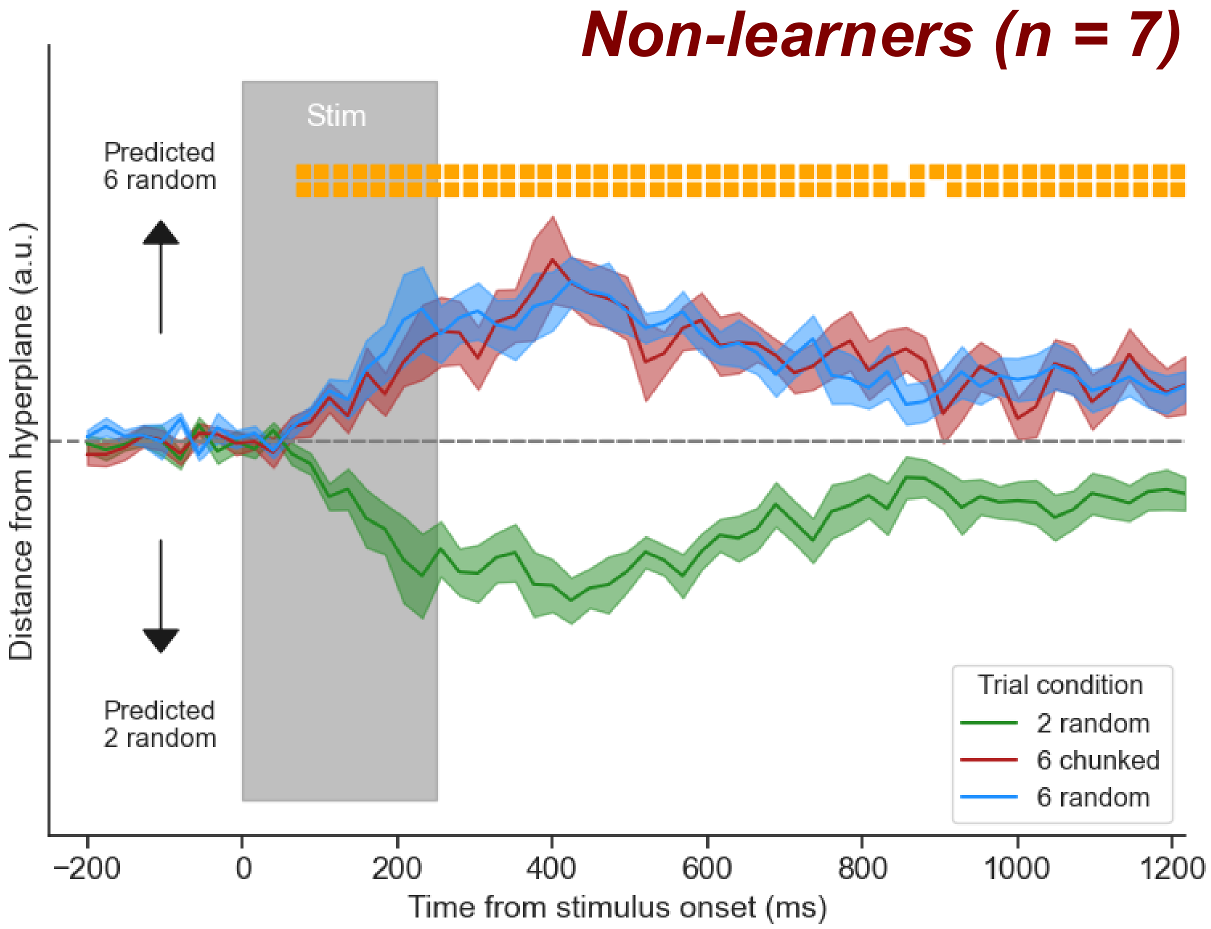


Training Results

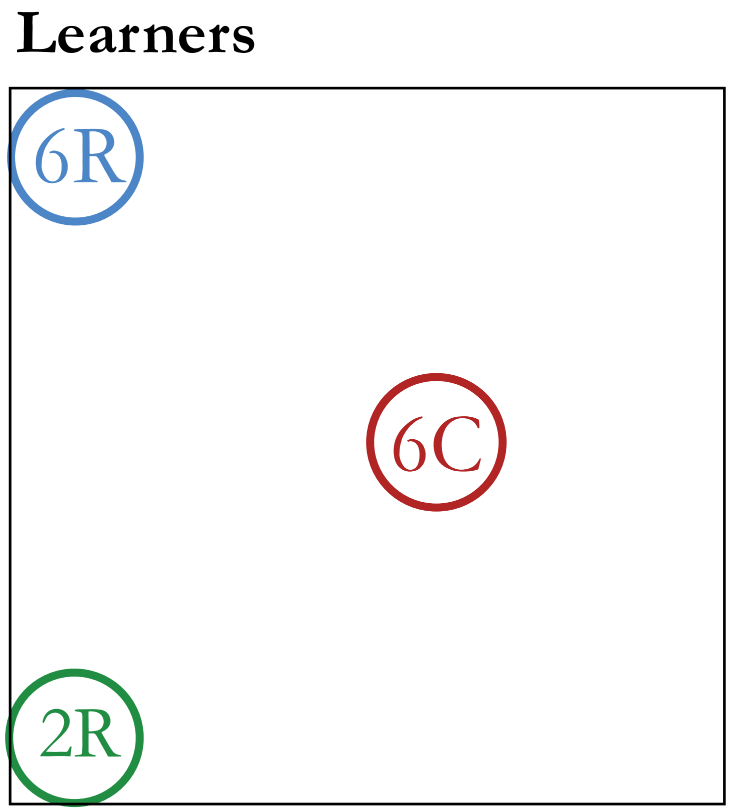


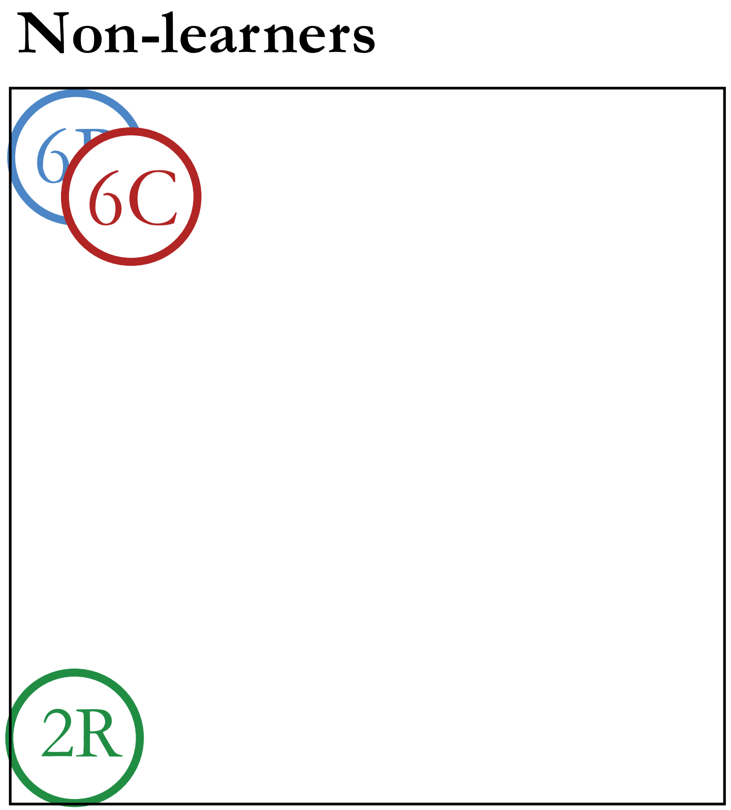
Learning changes the representational geometry





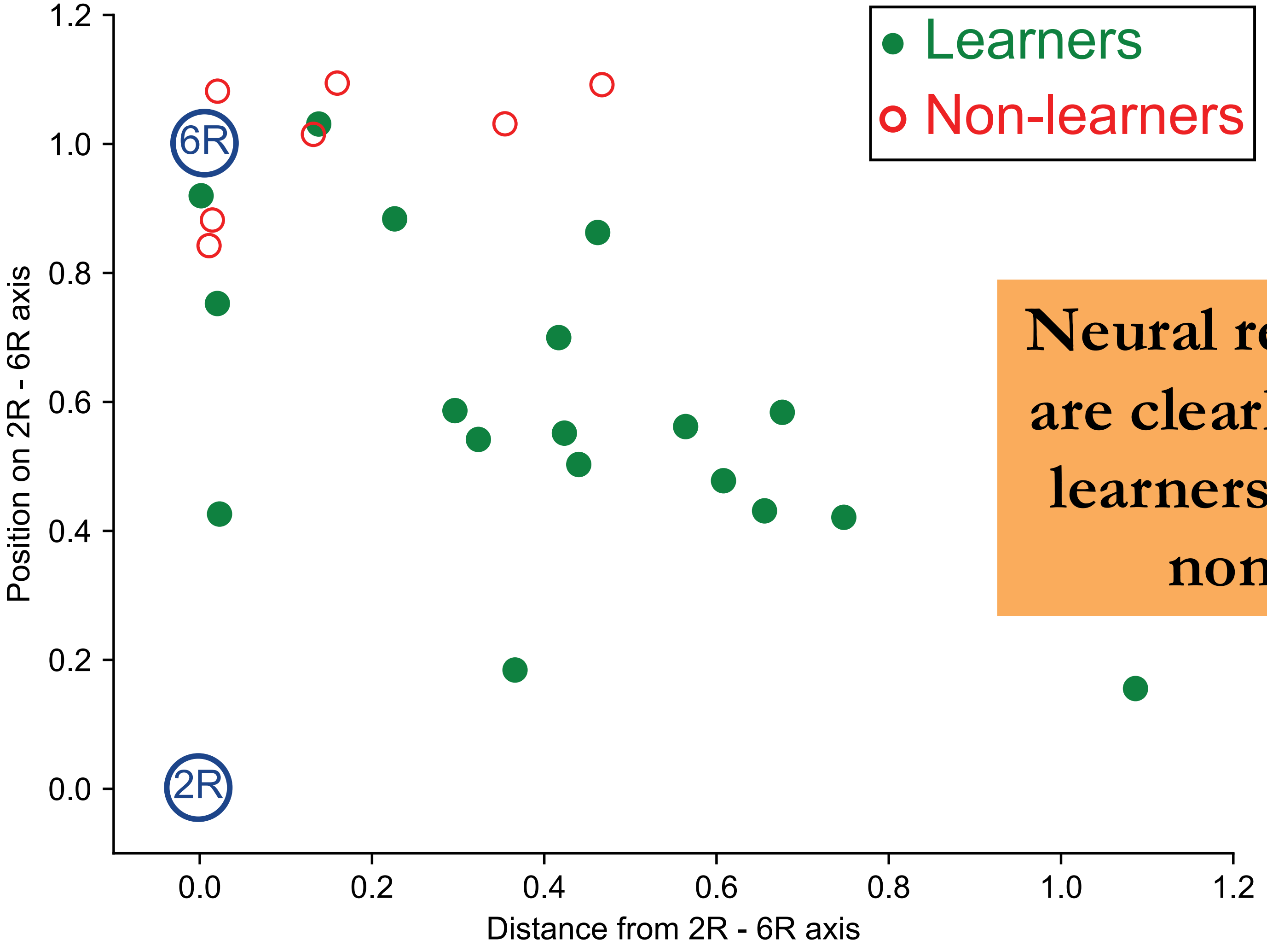
Learning changes the representational geometry





Learners likely are using *different* representations and processes to non-learners and that is reflected in their neural activity.

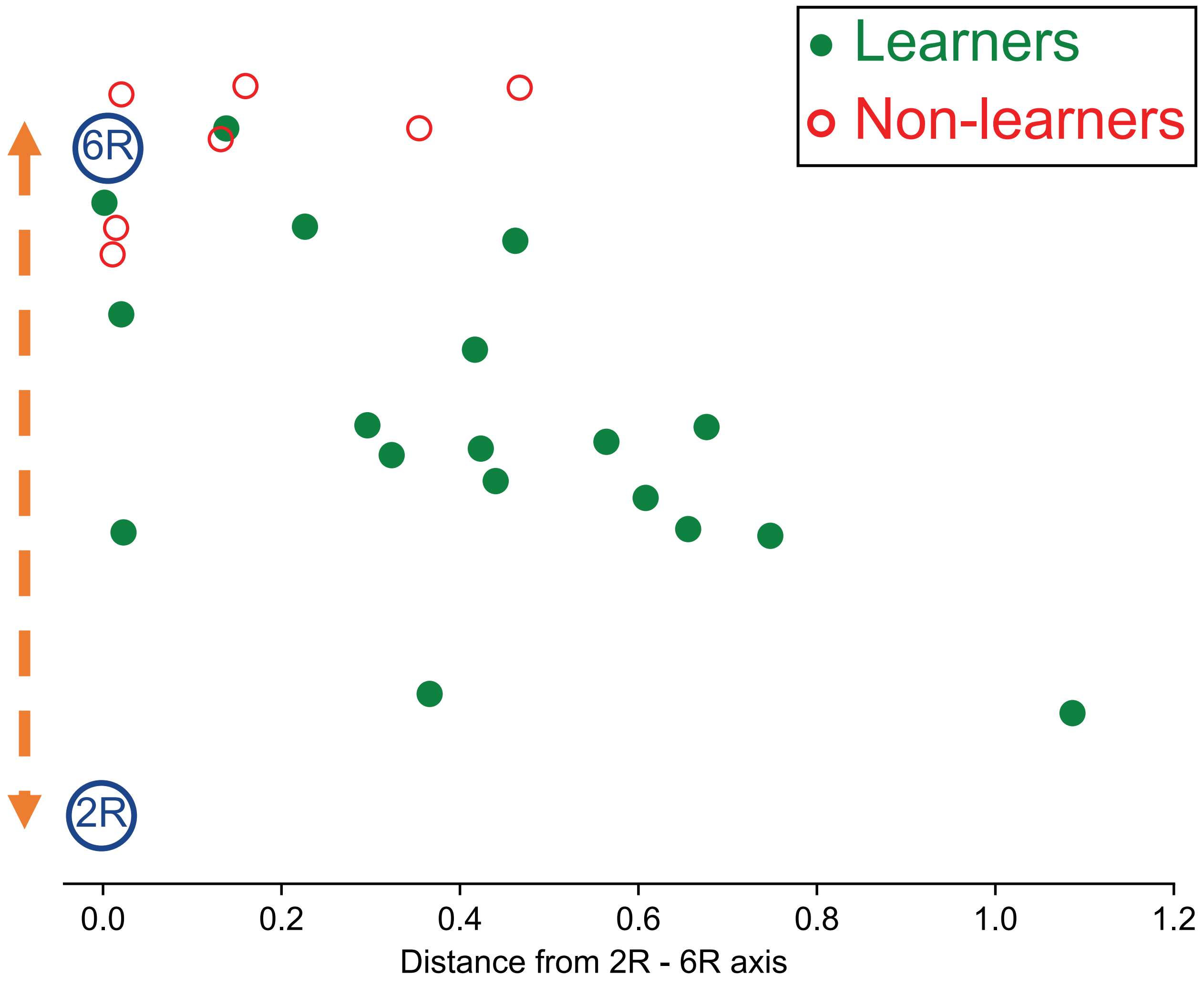
Multidimensional scaling on each subject



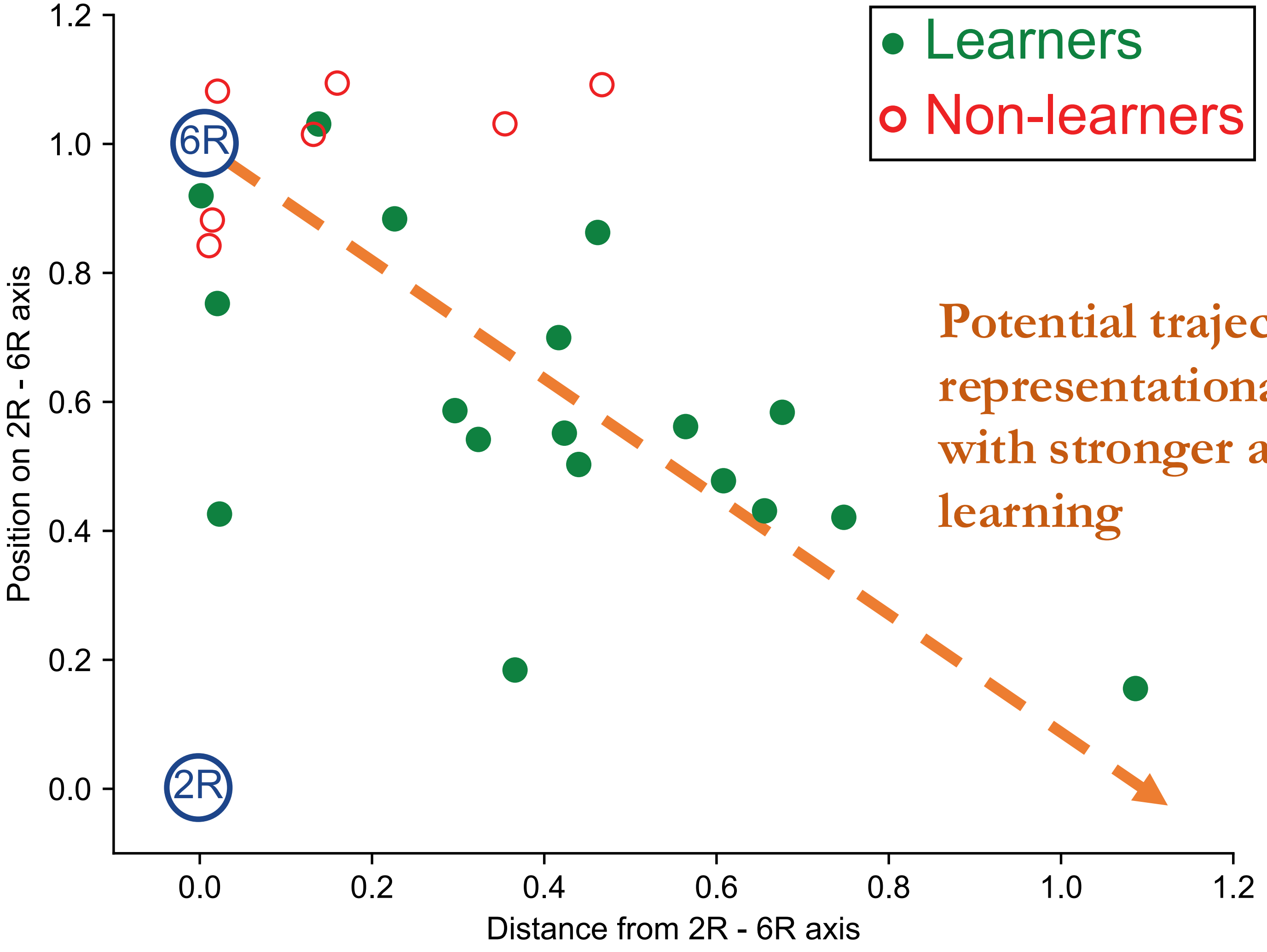
Neural representations are clearly different for learners compared to non-learners

Multidimensional scaling on each subject

We assume increased neural similarity towards 2R (and dissimilarity to 6R) suggests a reduction in working memory load

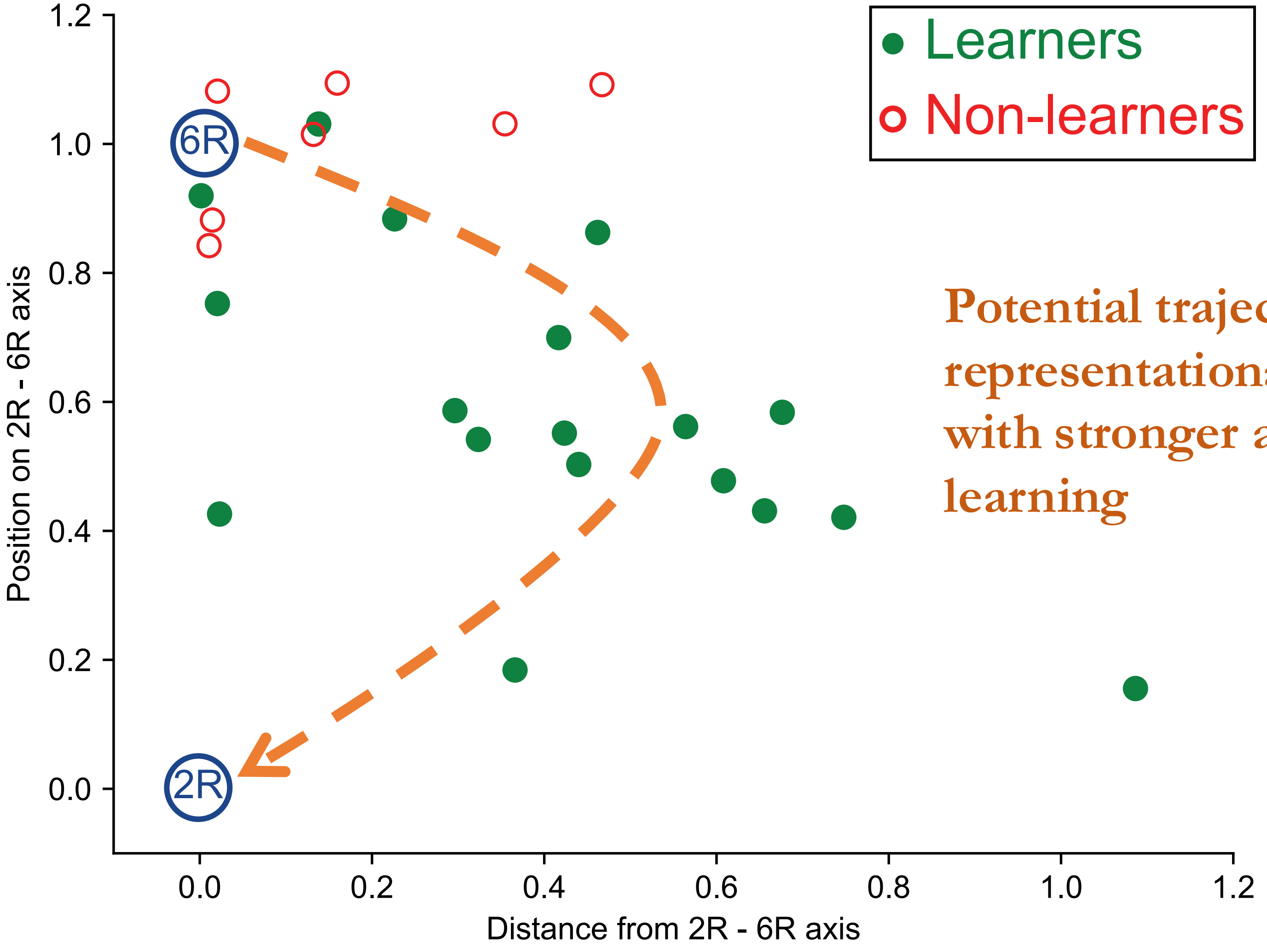


Multidimensional scaling on each subject



Potential trajectory of representational similarity with stronger associative learning

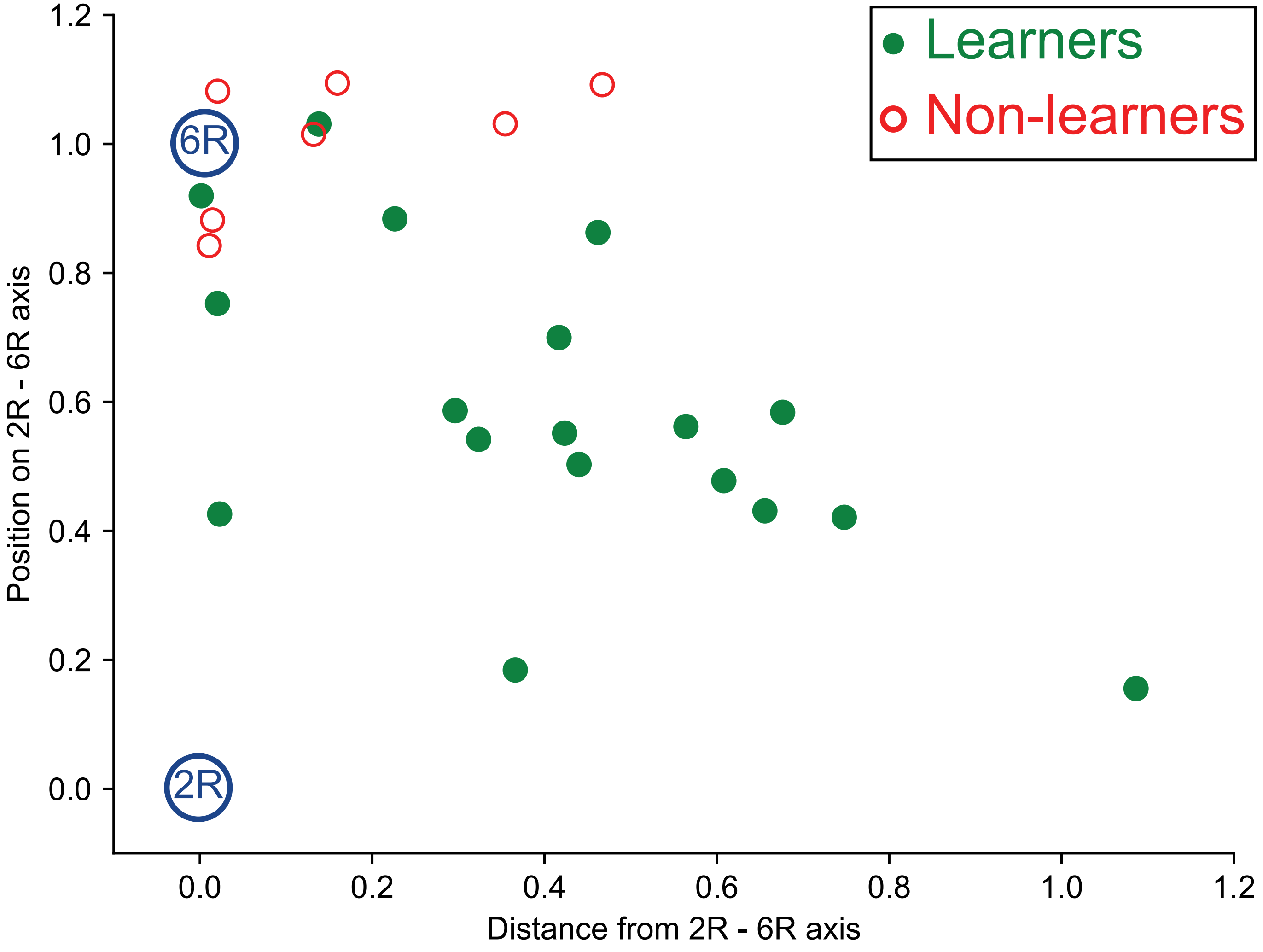
Multidimensional scaling on each subject



Potential trajectory of representational similarity with stronger associative learning

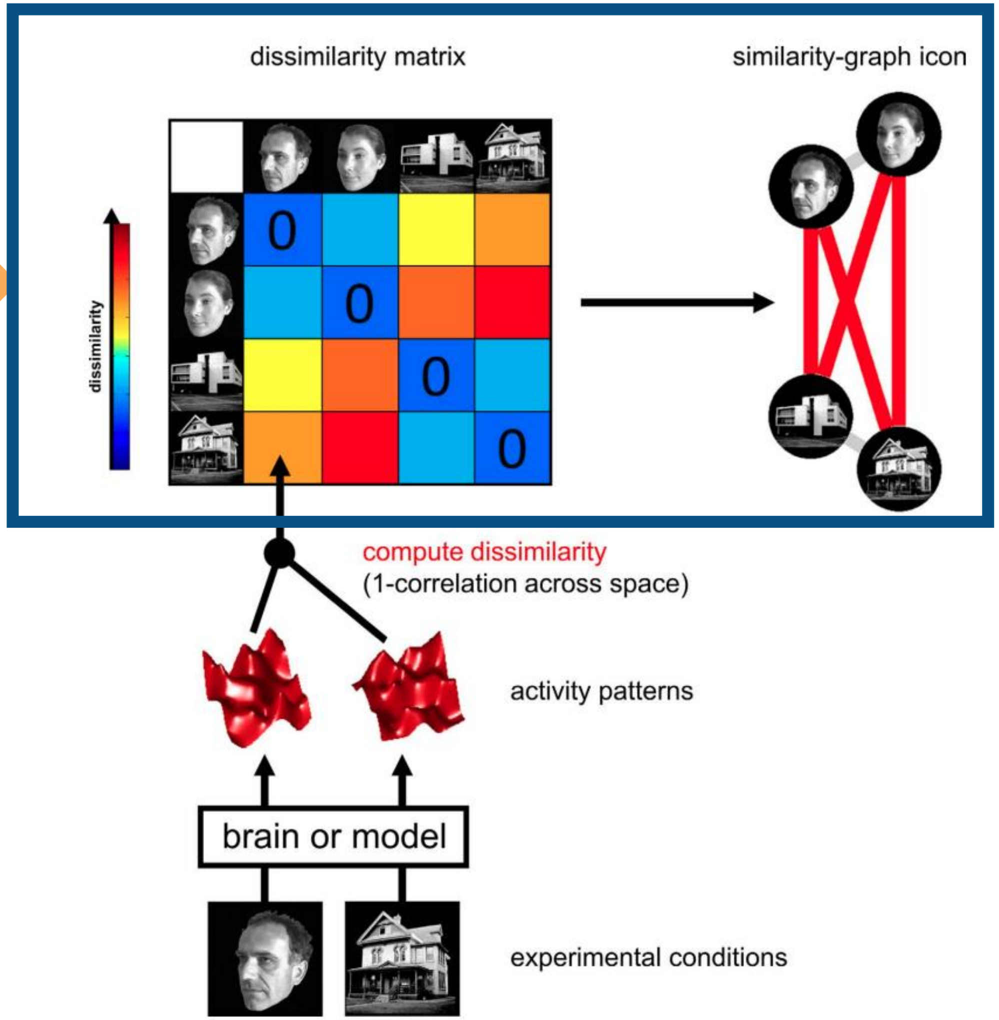
Multidimensional scaling on each subject

We ought to have a cognitive model that **explains** the neural representation and **make specific predictions** of the representational similarity so that it is **falsifiable.**



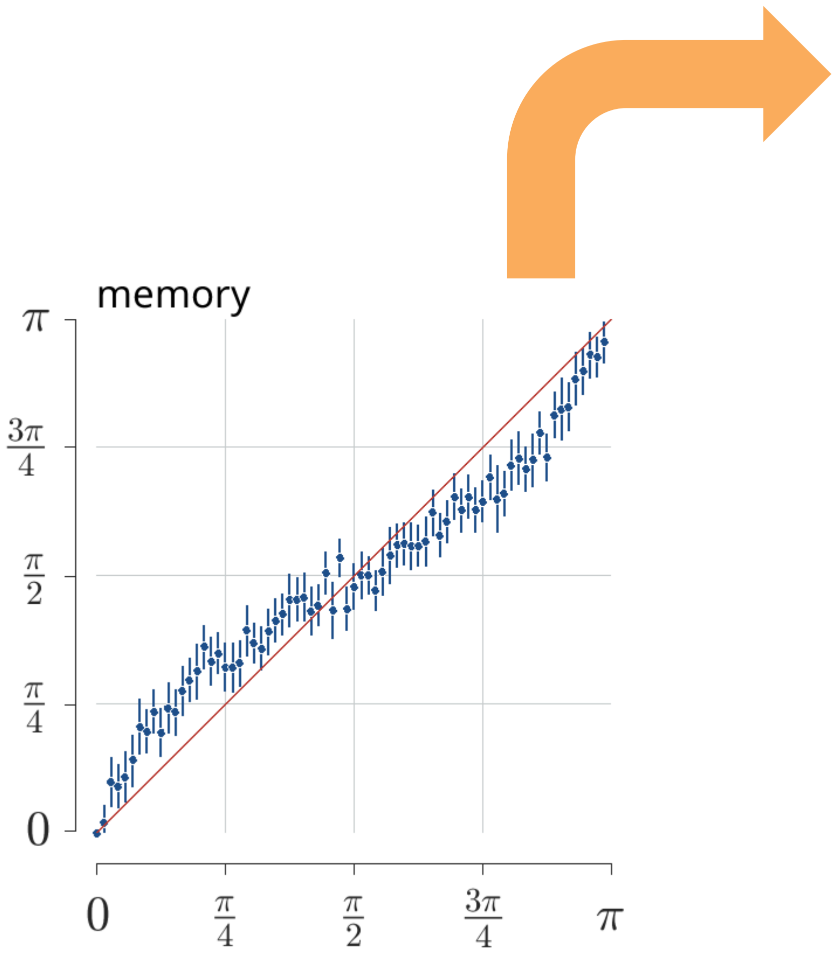
Representational similarity analysis

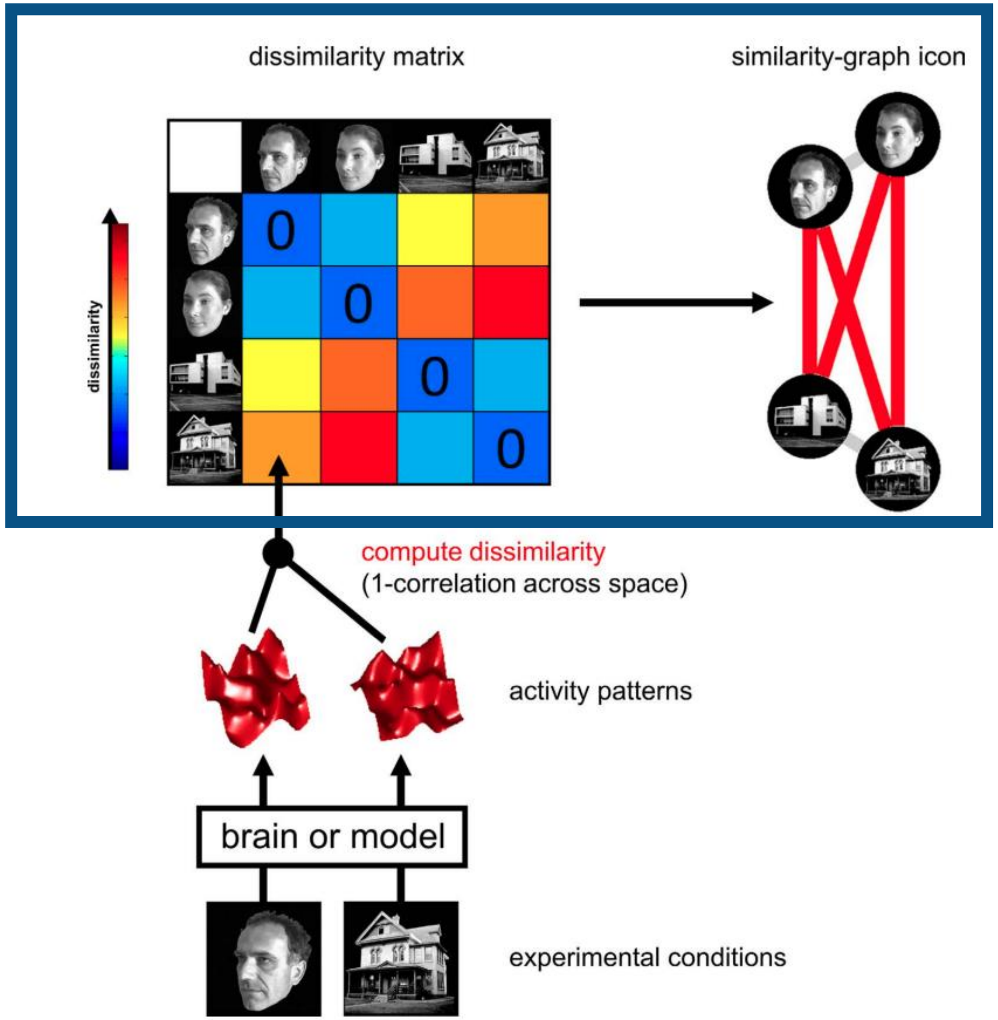
We need **cognitive models** to predict the observed representational (dis)similarity



Kriegeskorte, Mur, & Bandettini (2008). https://doi.org/10.3389/neuro.06.004.2008

Representational alignment





Kriegeskorte, Mur, & Bandettini (2008). https://doi.org/10.3389/neuro.06.004.2008

Where to from here?

Cognitive psychologists and neuroscientists are trying to integrate their findings and models (a.k.a. **representational alignment**)

But I think this will **not** be achieved without clear constraints on how researchers wield the concept of representation.

Sucholutsky, I., Muttenthaler, L., Weller, A., Peng, A., Bobu, A., Kim, B., ... & Griffiths, T. L. (2023). Getting aligned on representational alignment. *arXiv preprint arXiv:2310.13018*.

Some (potential) desiderata for representations

- Have a **causal role** in the cognitive behaviour (i.e. if the representation fails, this produces a failure in behaviour).
- Be dynamic or adaptive when required by the task-at-hand.
- Be constrained by biophysical principles (i.e. energy-efficient, anatomical structures)

O'Brien, G., & Opie, J. (2004). Notes toward a structuralist theory of mental representation. In *Representation in mind* (pp. 1-20). Elsevier.

Will aligning representational geometries be helpful?

I don't know!

Like how similarity ought not to be the assumed basis for all cognition (though convenient), judging similarity of representations may be a helpful but not verdicial basis.

My hope is that direct connection of representational geometries in behaviour and in the brain produces **some progress…**

O'Brien, G., & Opie, J. (2004). Notes toward a structuralist theory of mental representation. In *Representation in mind* (pp. 1-20). Elsevier.

Thank you!

Preprint with

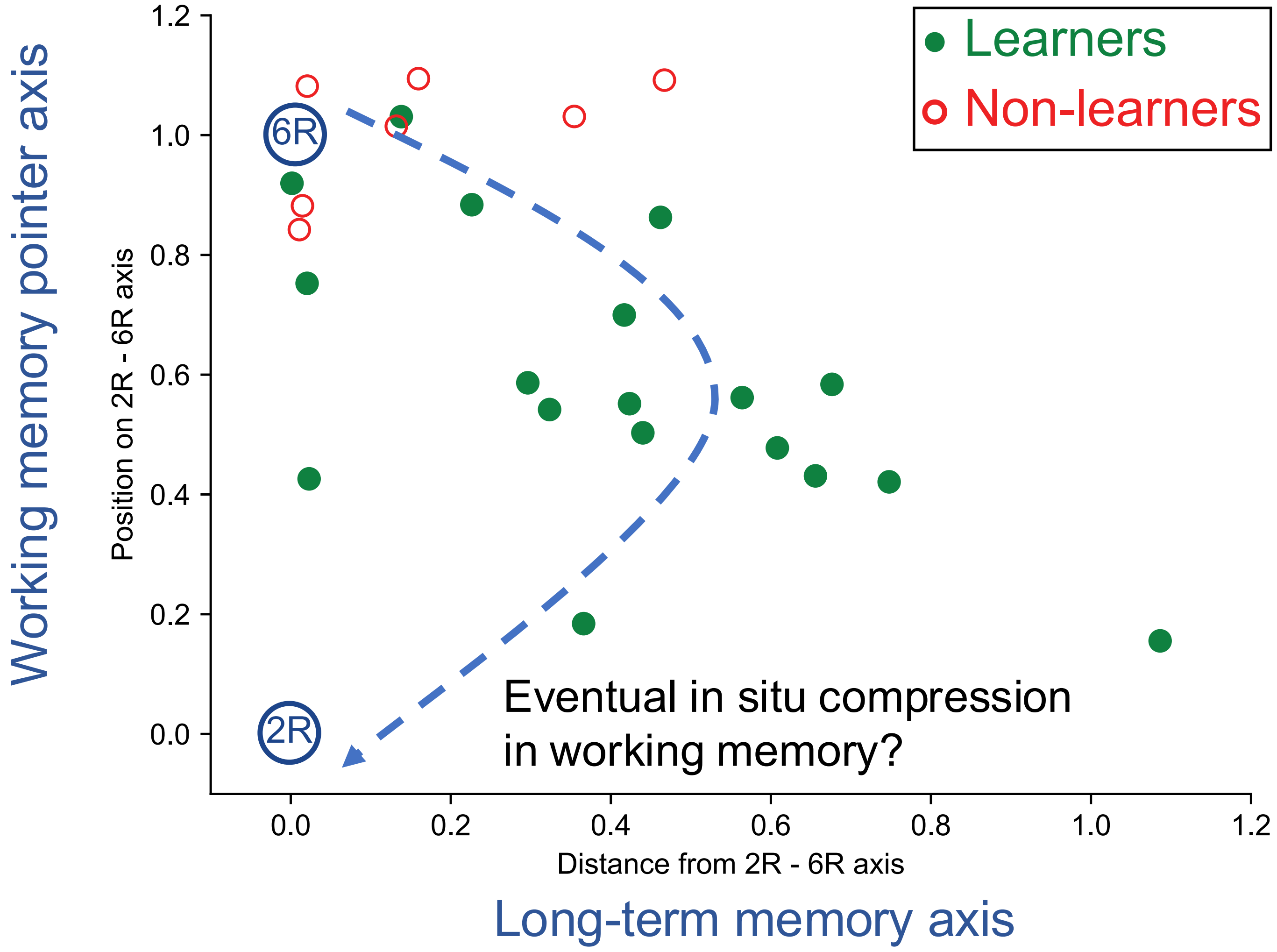
william.ngiam@adelaide.edu.au

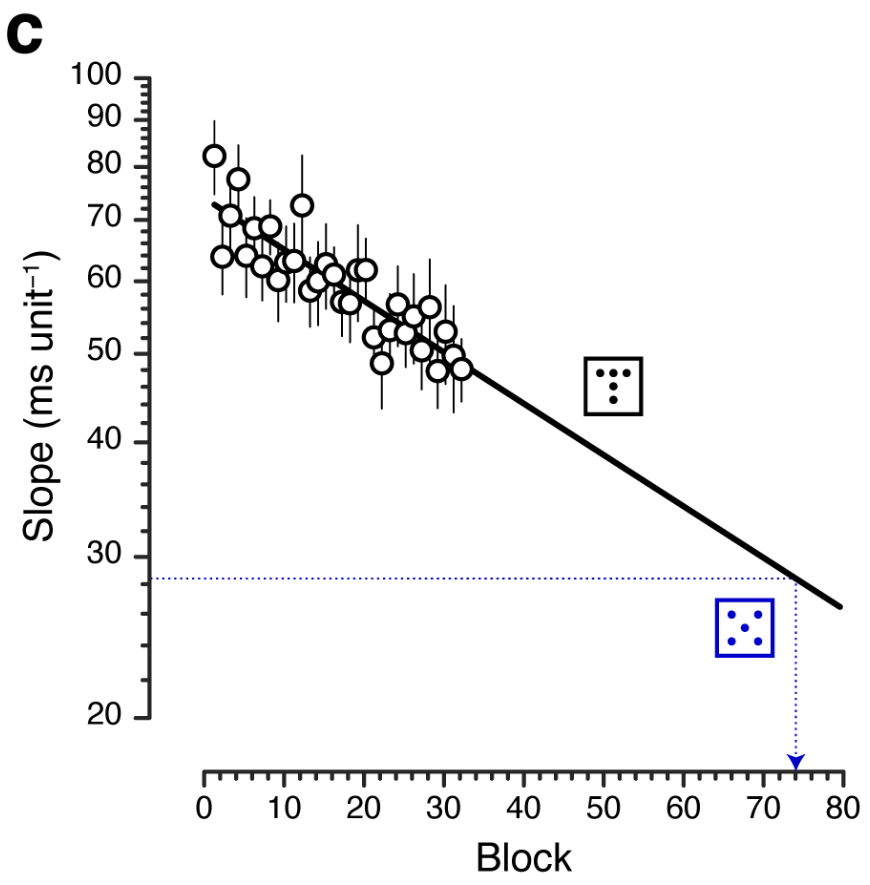
https://palm-lab.github.io

@williamngiam.github.io

Adelaide University

Bonus slide: why does learning fully reduce load // learned condition not cross the hyperplane?



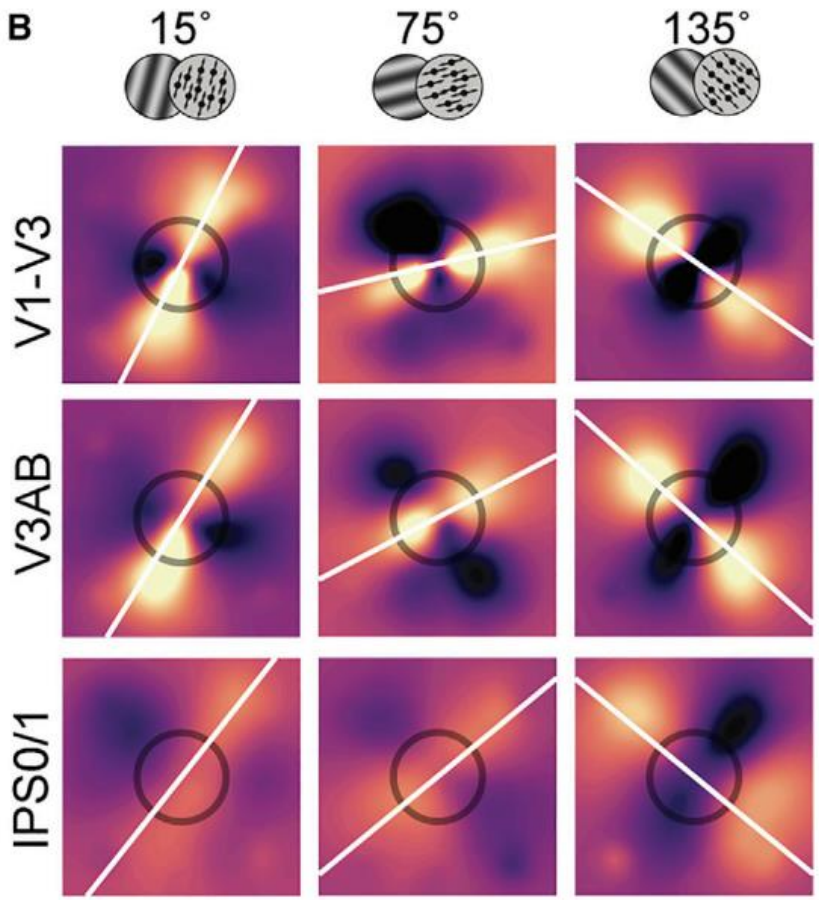


https://osf.io/preprints/psyarxiv/gh5ps

Representations to *in vivo* neuroscientists

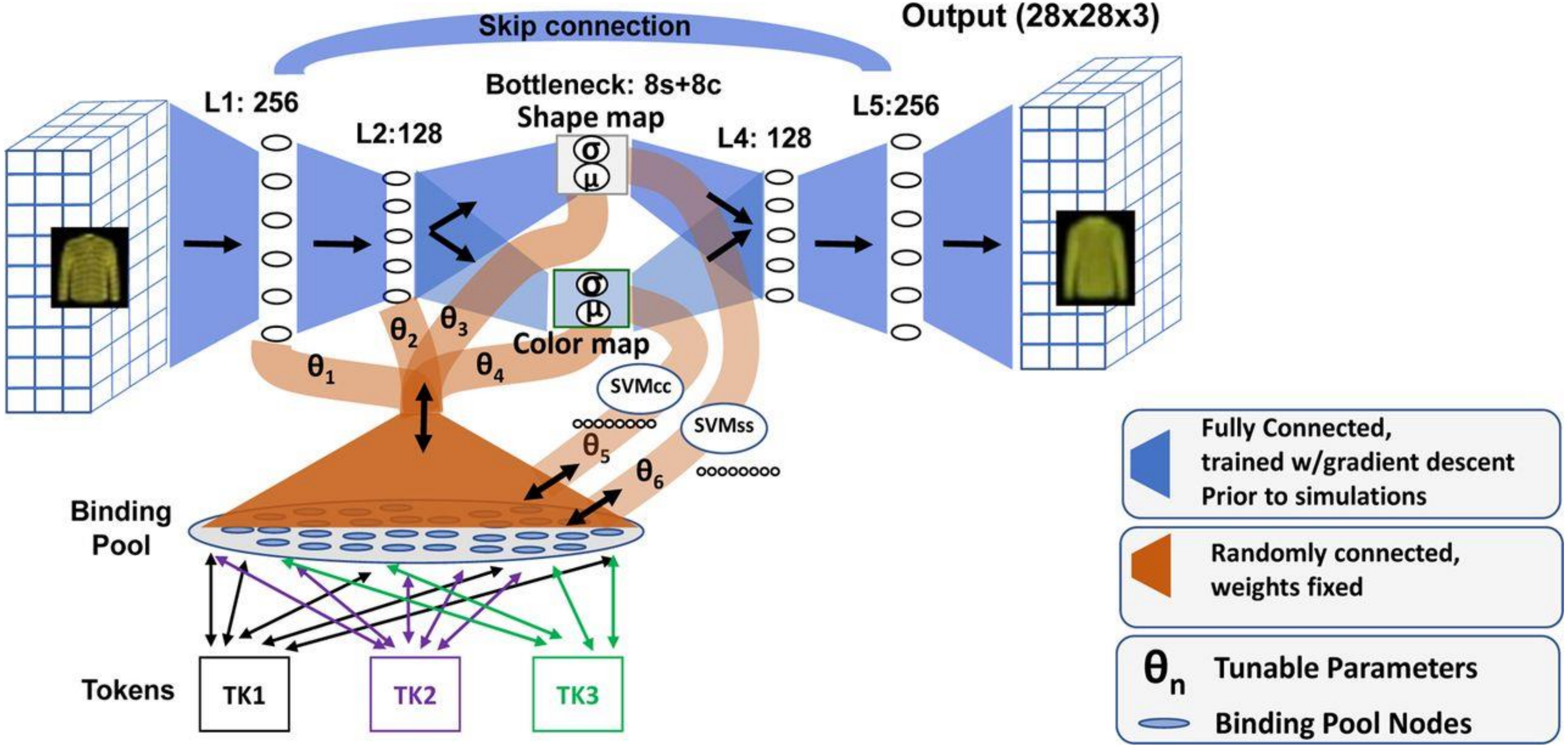
The neuroscientists think of the representations as **activity in a population of neurons** or **across a network of brain regions**.

But this does not map neatly to our psychological models. When does perception become working memory?



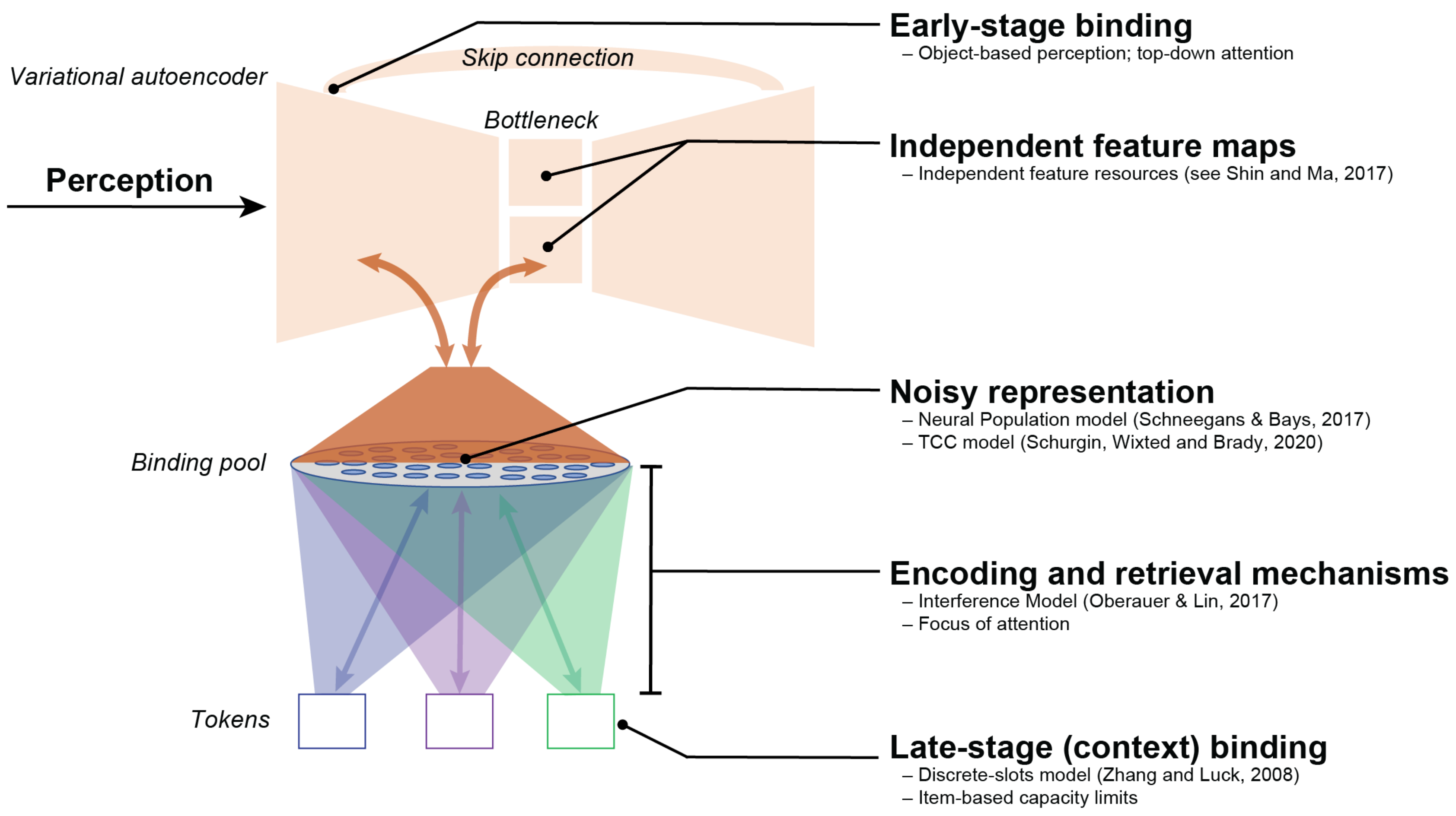
Kwak and Curtis (2022) https://doi.org/10.1016/j.neuron.2022.03.016

Representations to *in silico* neuroscientists



Hedayati, O'Donnell, & Wyble (2022). https://doi.org/10.1038/s41562-021-01264-9

An attempt to map everything...

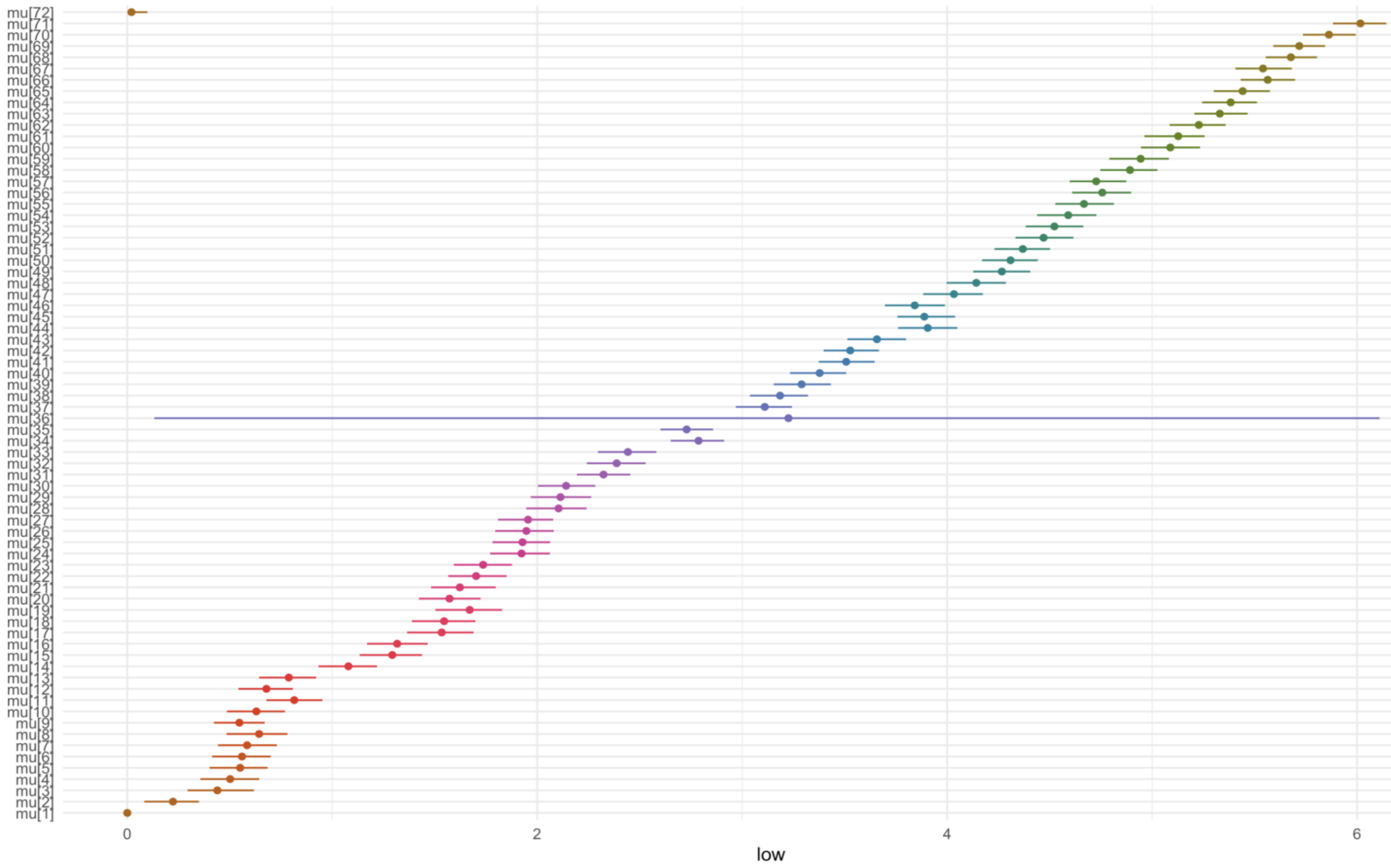


Hedayati, O'Donnell, & Wyble (2022). https://doi.org/10.1038/s41562-021-01264-9
Ngiam (2024). https://doi.org/10.3758/s13423-023-02356-5

Bonus slide – latent representation of colour

We applied the same model to the similarity judgment task that used **colour stimuli**. (**WIP**)

Each colour in the plot is what was shown in the experiment.



Bonus slide – latent representation of colour