Correcting bias in visual working memory researchers: *Steps towards an integrative framework*

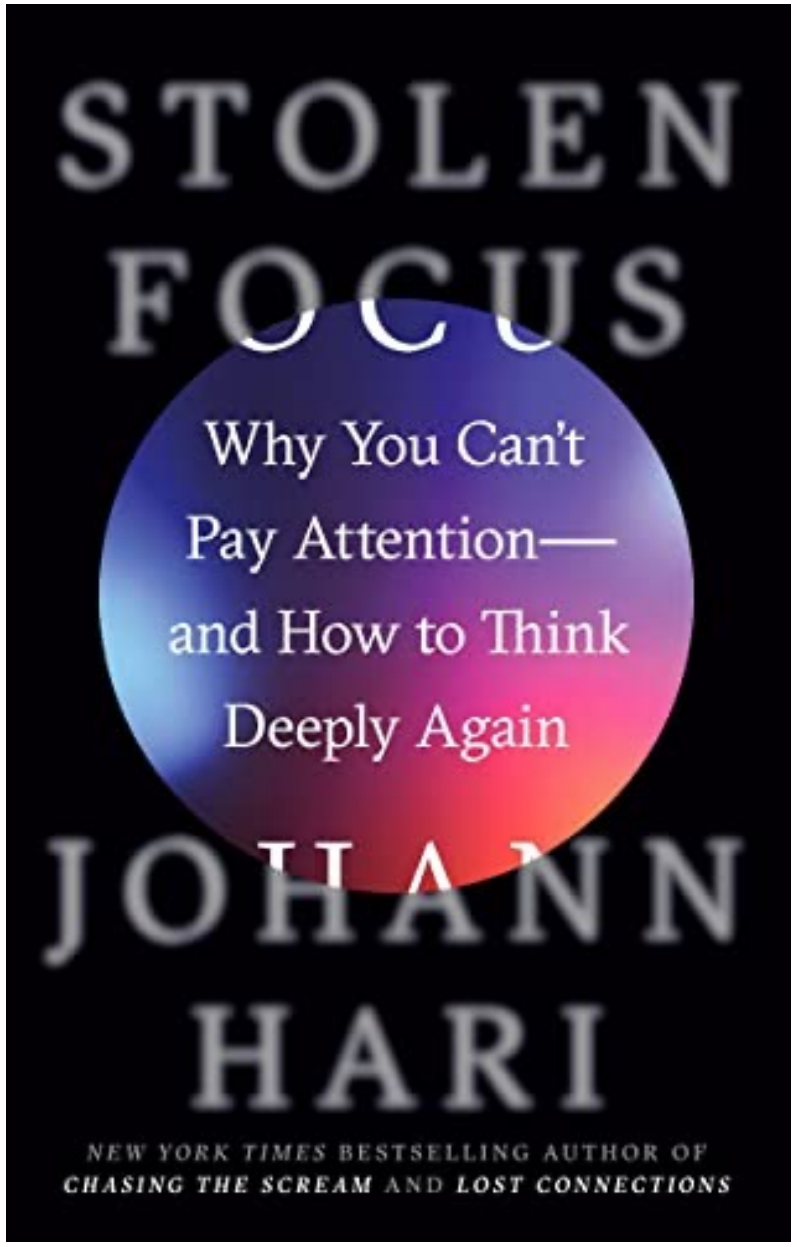
Dr William XQ Ngiam

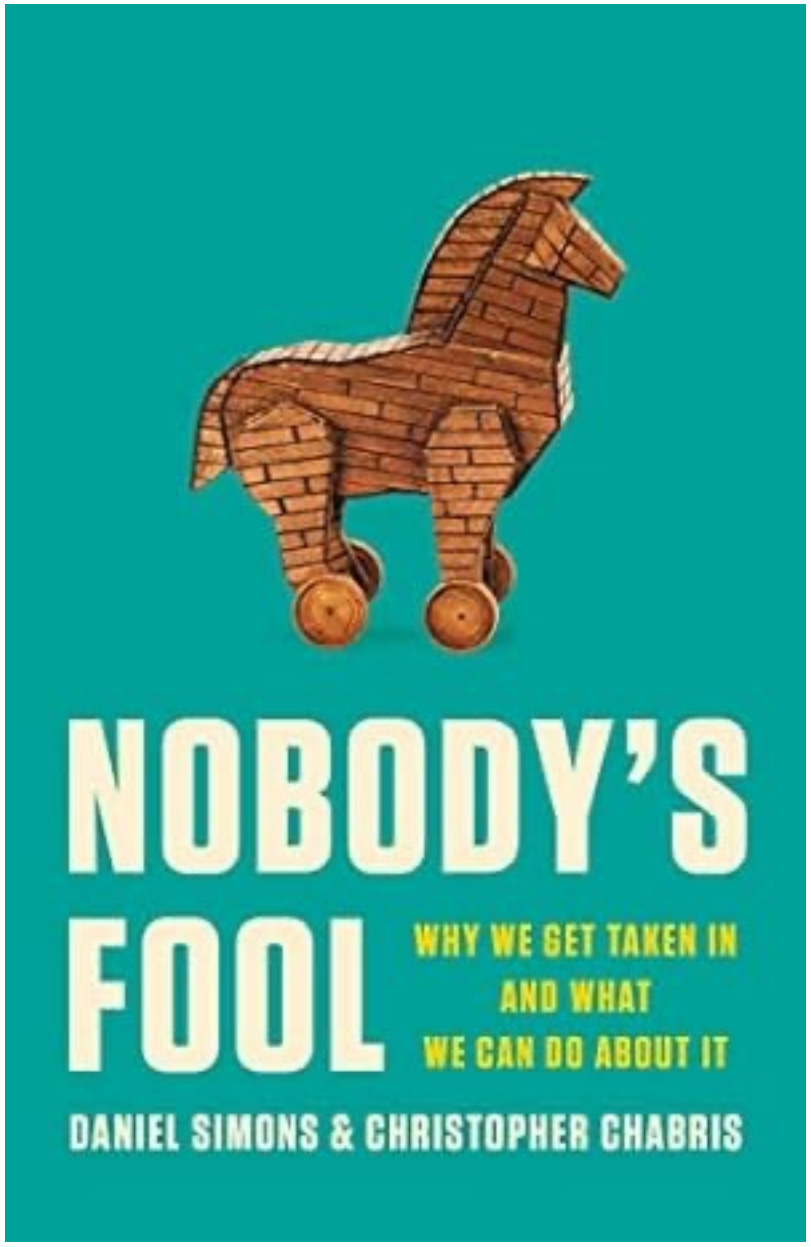
Postdoctoral Researcher in the Awh/Vogel Lab

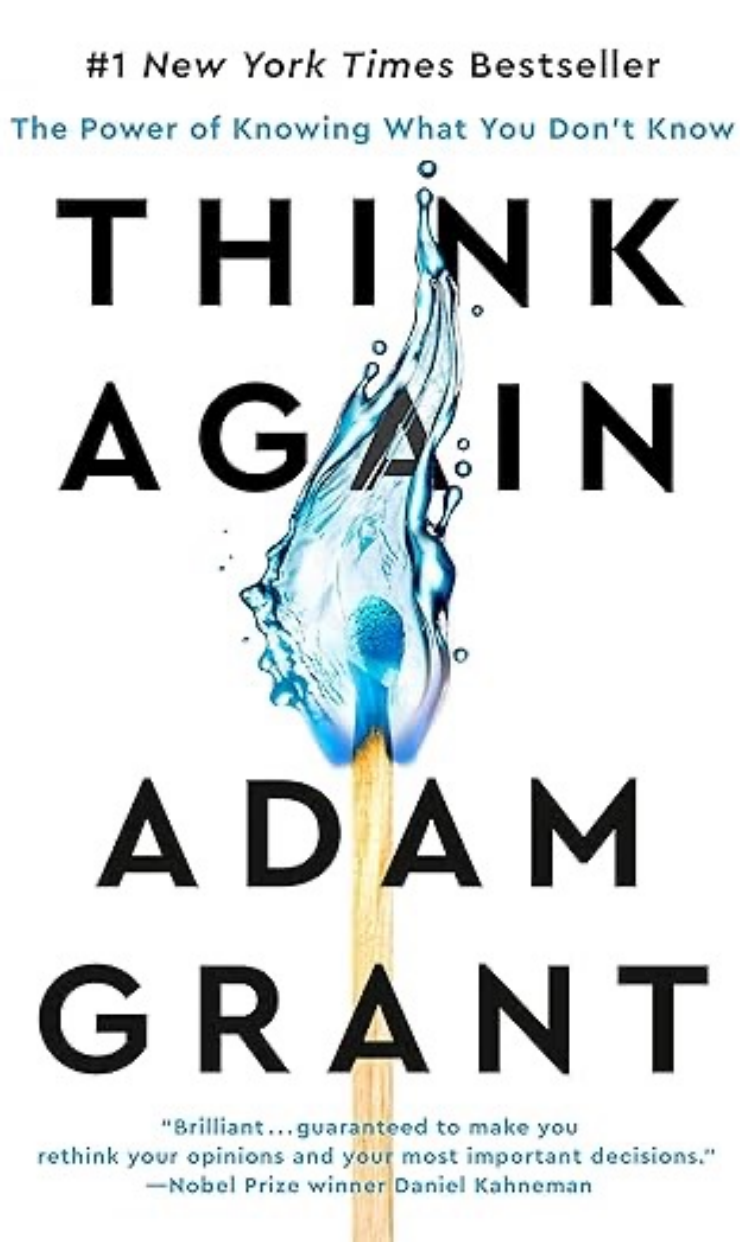
THE UNIVERSITY OF CHICAGO

INSTITUTE FOR MIND AND BIOLOGY
THE UNIVERSITY OF CHICAGO

for the Ebbinghaus Empire Series, University of Toronto


STOLEN FOCUS
Why You Can't Pay Attention— and How to Think Deeply Again
JOHANN HARI
NEW YORK TIMES BESTSELLING AUTHOR OF
CHASING THE SCREAM AND LOST CONNECTIONS


NOBODY'S FOOL
WHY WE GET TAKEN IN AND WHAT WE CAN DO ABOUT IT
DANIEL SIMONS & CHRISTOPHER CHABRIS


#1 New York Times Bestseller
The Power of Knowing What You Don't Know
THINK AGAIN
ADAM GRANT
"Brilliant... guaranteed to make you rethink your opinions and your most important decisions."
—Nobel Prize winner Daniel Kahneman

1. What are scientists thinking?

2. What is visual working memory?
    a) *Guess bands*
    b) *Conjunction whole-report*

3. How do we rethink visual working memory?

What are scientists thinking?

Exponential growth of scientific publications

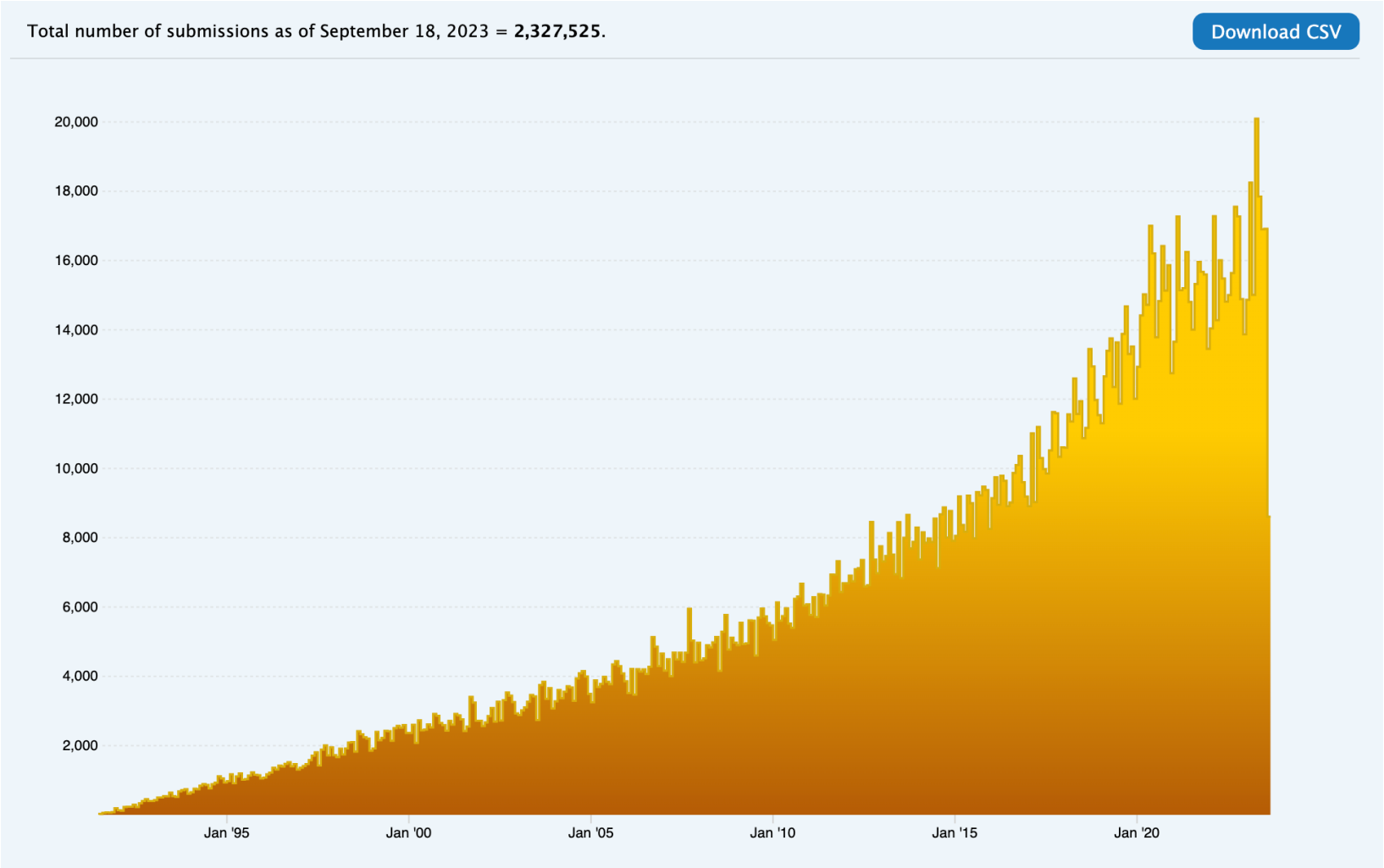


Figure taken from arxiv.org on the number of submissions over time. https://arxiv.org/stats/monthly_submissions

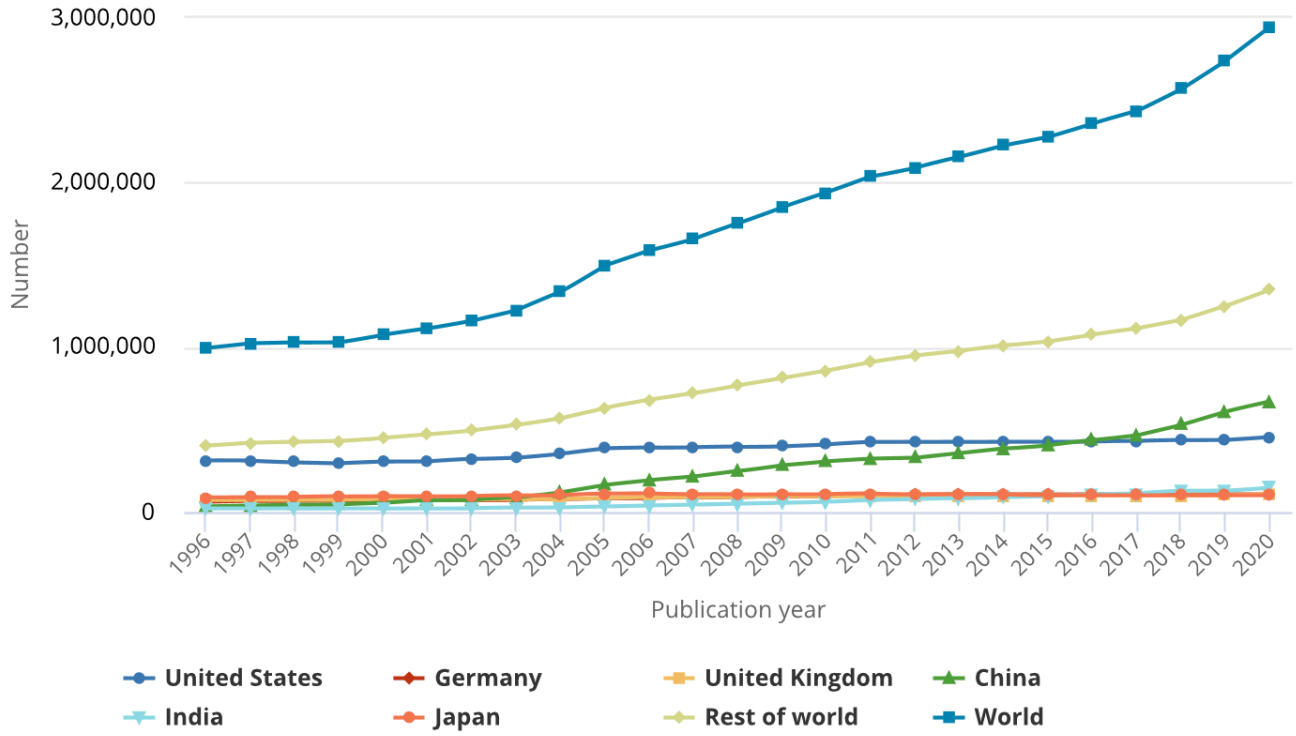
Exponential growth of scientific publications

- Estimated to have reached 2.9 million articles in 2020 (National Science Board, National Science Foundation)
- Increasing by approximately 4% each year (Pan, Petersen, Pammolli and Fortunato, 2016)

National Center for Science and Engineering Statistics | NSB-2021-4

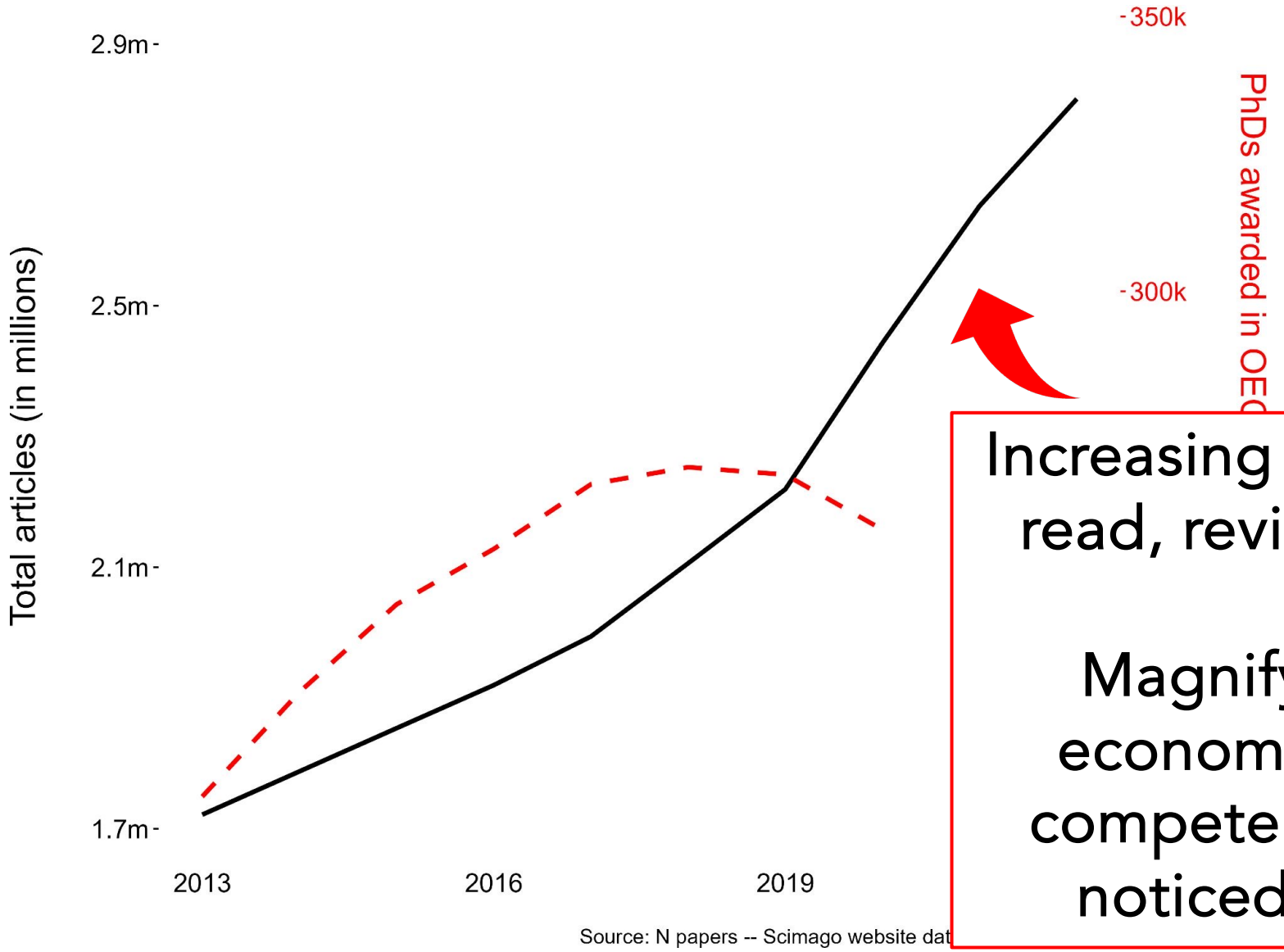
Figure PBS-2

S&E articles, by selected region, country, or economy and rest of world: 1996–2020



Review by National Center for Science and Engineering Statistics. https://ncses.nsf.gov/pubs/nsb20214/publication-output-by-country-region-or-economy-and-scientific-field

Pan, R. K., Petersen, A. M., Pammolli, F., & Fortunato, S. (2018). The memory of science: Inflation, myopia, and the knowledge network. *Journal of Informetrics*, *12*(3), 656-678. https://arxiv.org/abs/1607.05606



Increasing strain on scientists to read, review and co-ordinate.

Magnifying an "attention economy" where scientists compete for their work to be noticed and have impact.

Figure copied from https://bsky.app/profile/hansonmark.bsky.social/post/3kajeqzv3nt2b

Hanson, Barreiro, Crosetto and Brockington (2023). The strain on scientific publishing. *ArXiv.* https://arxiv.org/abs/2309.15884

The decline of negative results

- The proportion of papers reporting a positive result has been increasing f

Does more papers (mostly with positive findings) mean faster scientific progress?

I say not really.

- In the recent psychology literature, this proportion is estimated to be ~95% (Scheel, Schijen and Lakens, 2021)

Figure from Fanelli, D. (2012). Negative results are disappearing from most disciplines and countries. *Scientometrics*, *90*(3), 891-904.
Scheel, A. M., Schijen, M. R., & Lakens, D. (2021). An excess of positive results: Comparing the standard Psychology literature with Registered Reports. Advances in Methods and Practices in Psychological Science, 4(2), 25152459211007467.

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.

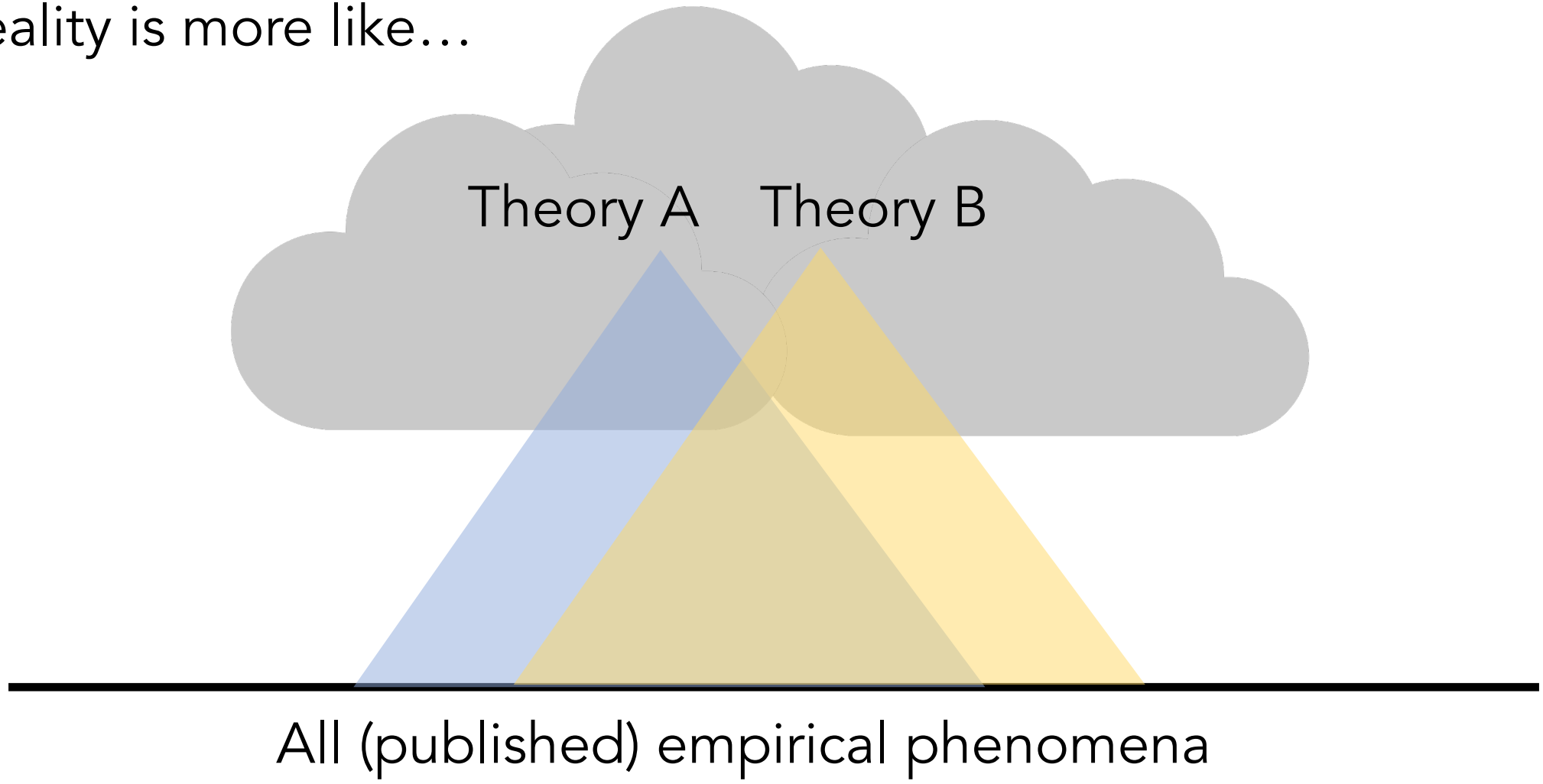
Playing *20 questions* with nature

- 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

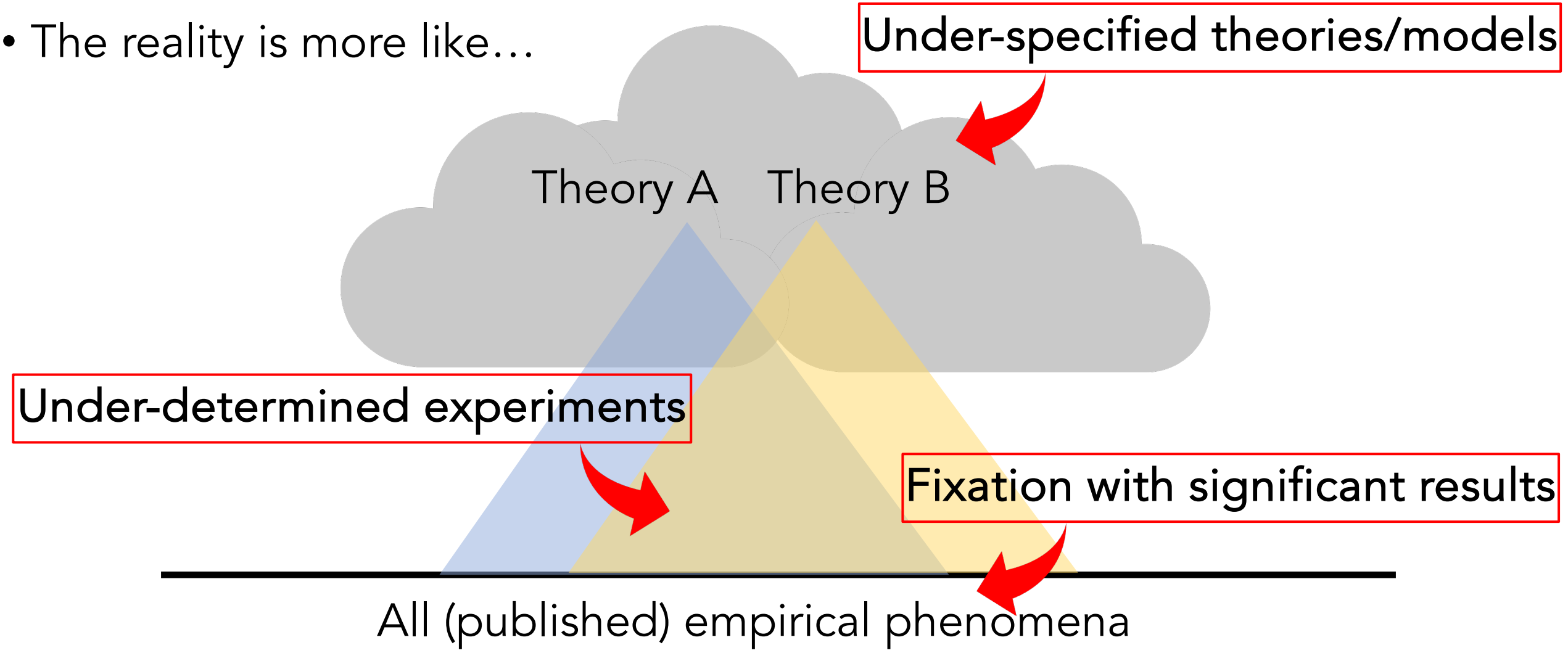
- The reality is more like…



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?

What *is* visual working memory?

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

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 visual imagery, psychological introspection

- What does it mean to *maintain* visual information?
- What details a state of *heightened accessibility*?

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

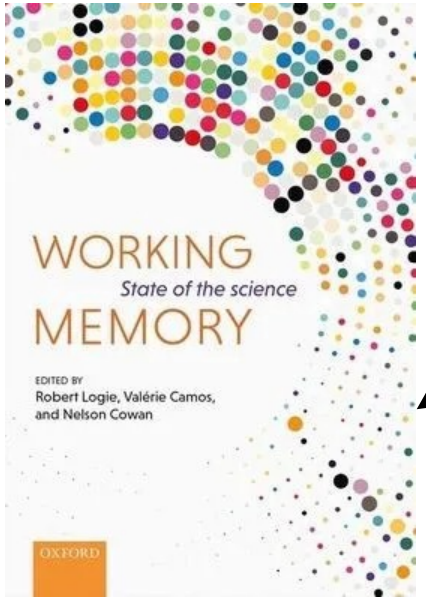
It has become clearer to me that a major source of confusion is that researchers use different definitions of the malleable and useful concept of WM. We do not seem to be converging on a common definition of the term. Others also have

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

How do we make progress if:

- There exist subtly different definitions
  - Due to different research questions, different methods, different measures, different contexts, etc.
- Theories (or models) attempt to explain all empirical phenomena related to ill-defined construct (overgeneralization)
- Models are underspecified such that empirical tests cannot be definitive
  - And these models may not reflect fundamental theoretical adversity

Guess bands

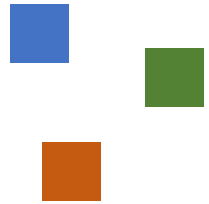
Joshua Foster

Kirsten Adam

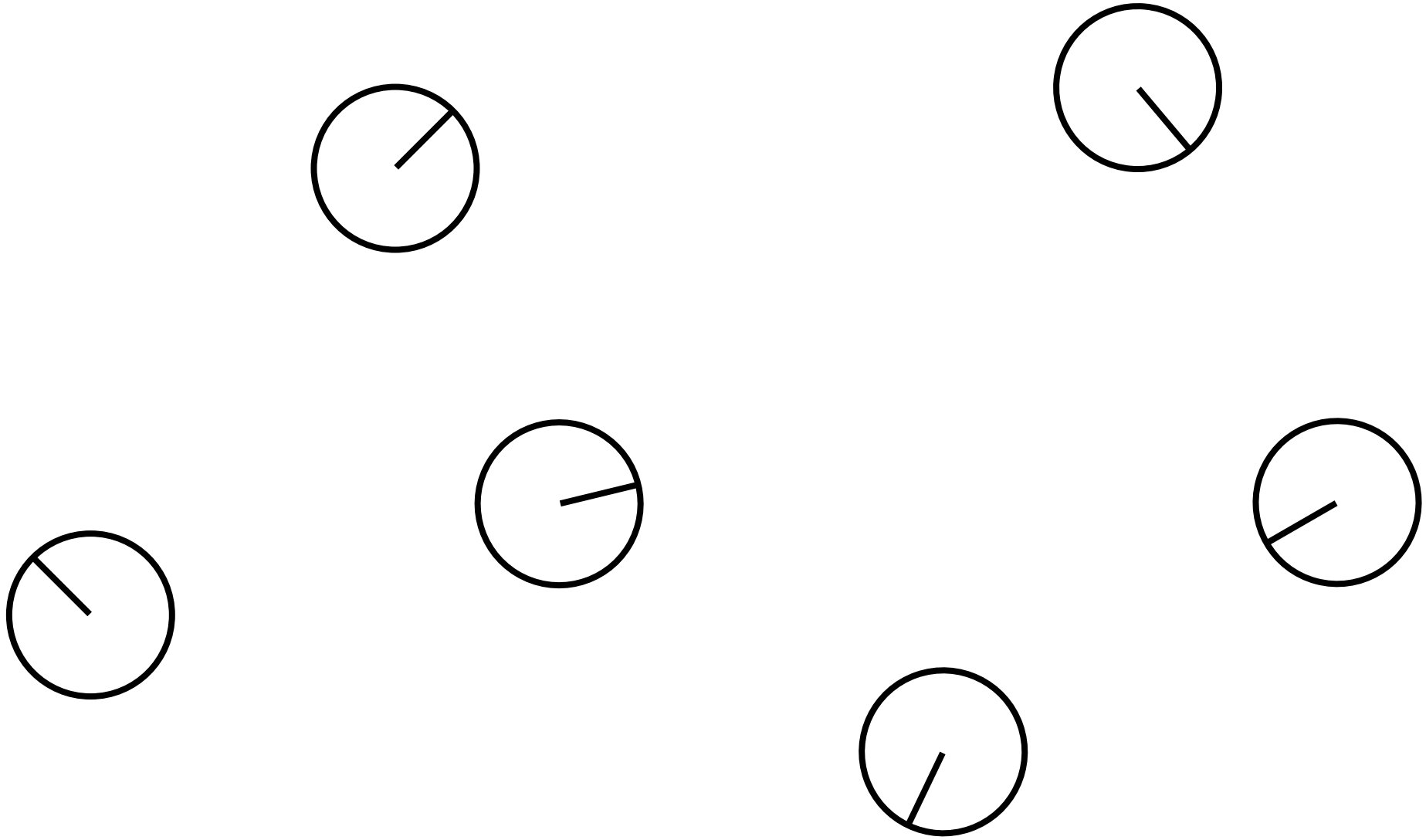
Ed Awh

What *is* visual working memory?

- Item-limit models (previously *slot models*)
  - Memory is contained to a few objects
  - There is no memory for objects beyond this capacity limit

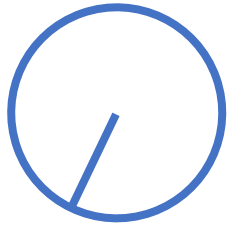


- Variable precision models (previously *flexible resource models*)
  - Memory is distributed across all items
  - There is flexible allocation of mnemonic resources to all items
    - More allocation of resources leads to a higher fidelity memory representation



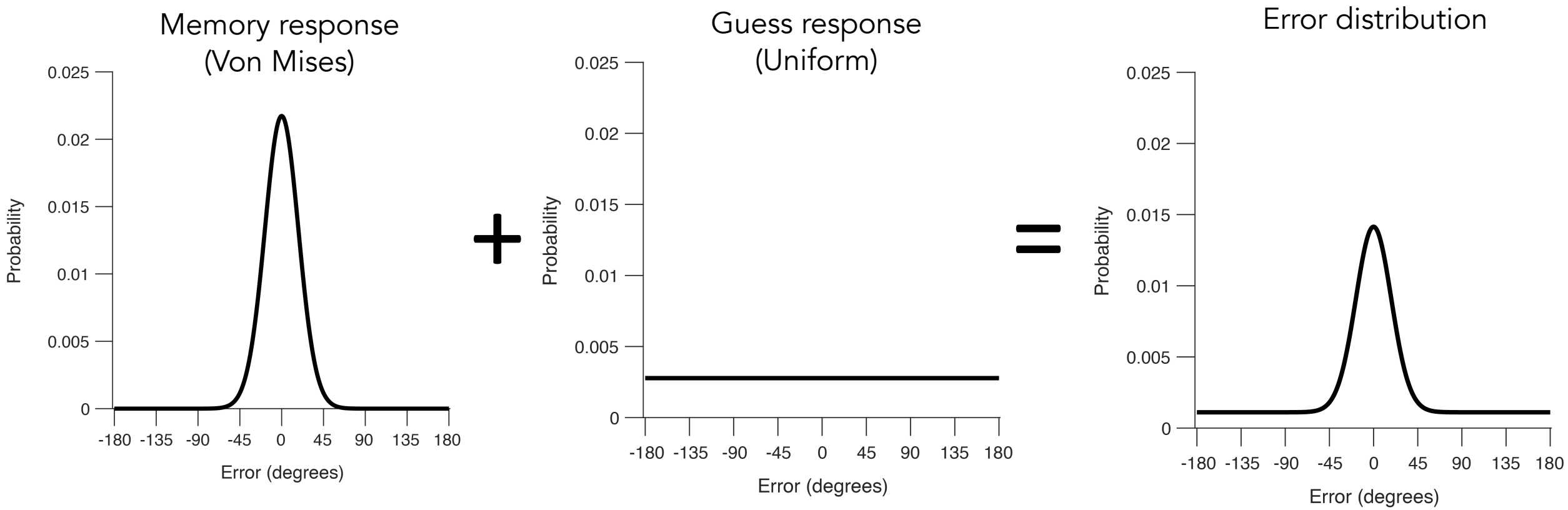
Three items are stored

But nothing for
the other items



Formal models

- Item-limit models (Zhang and Luck, 2008)



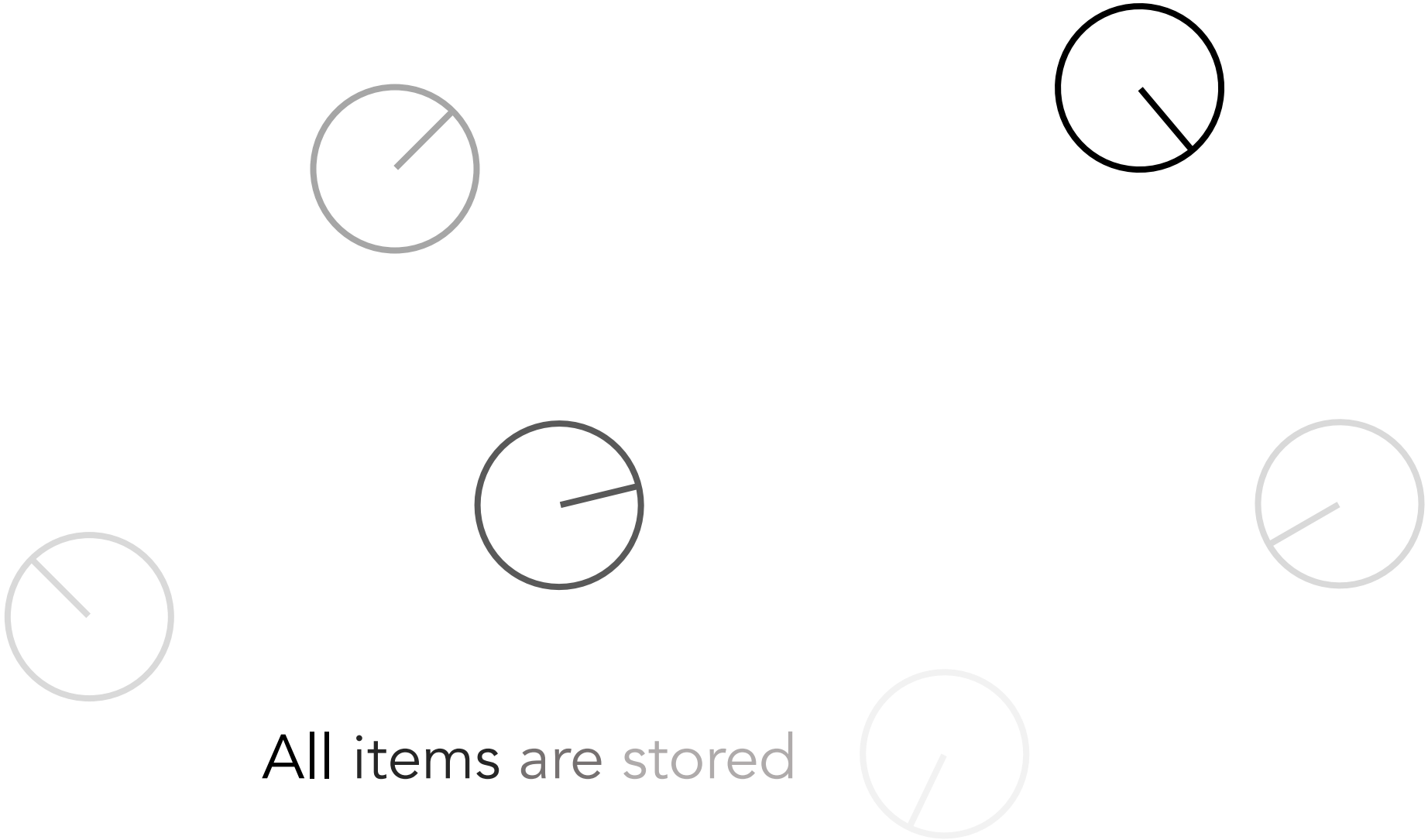
Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature*, *453*(7192), 233-235.

The competing models

- Item-limit models (previously *slot models*)
  - Memory is contained to a few objects
  - There is no memory for objects beyond this capacity limit

- Variable precision models (previously *flexible resource models*)
  - Memory is distributed across all items
  - There is flexible allocation of mnemonic resources to all items
    - More allocation of resources leads to a higher fidelity memory representation

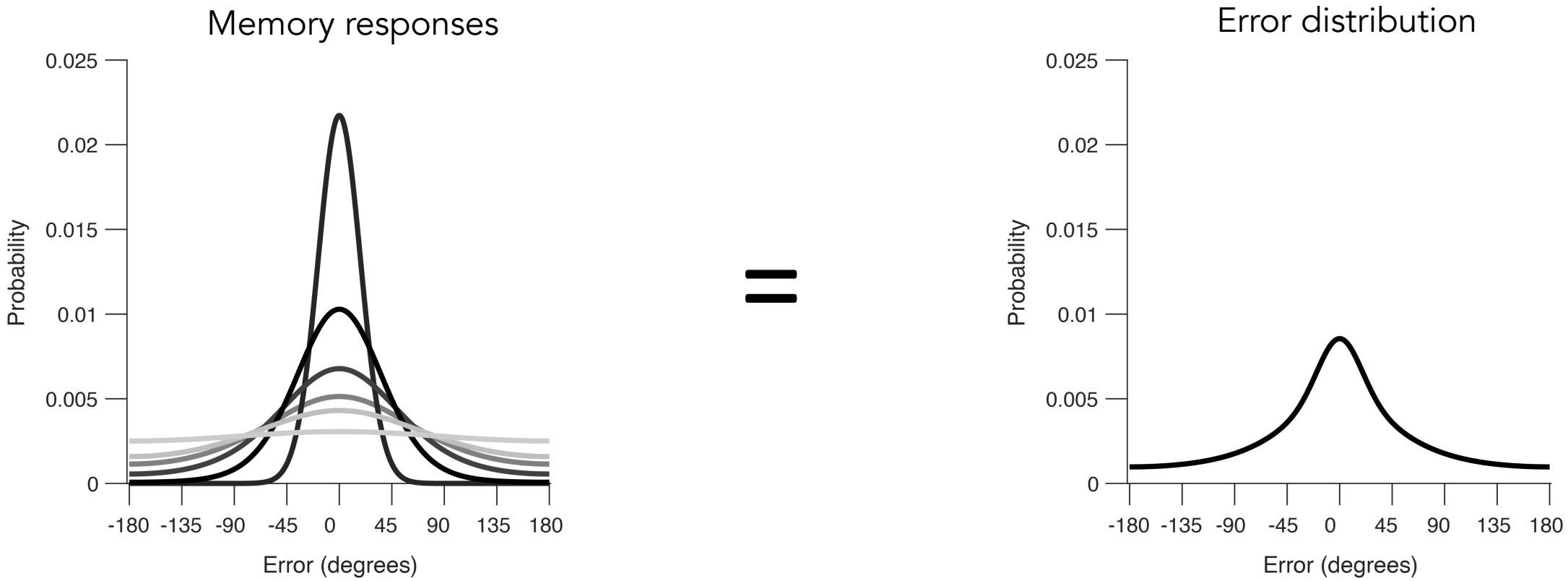
NB. An item limit is not mutually exclusive with a variable precision process (more on this later).



All items are stored

Formal models

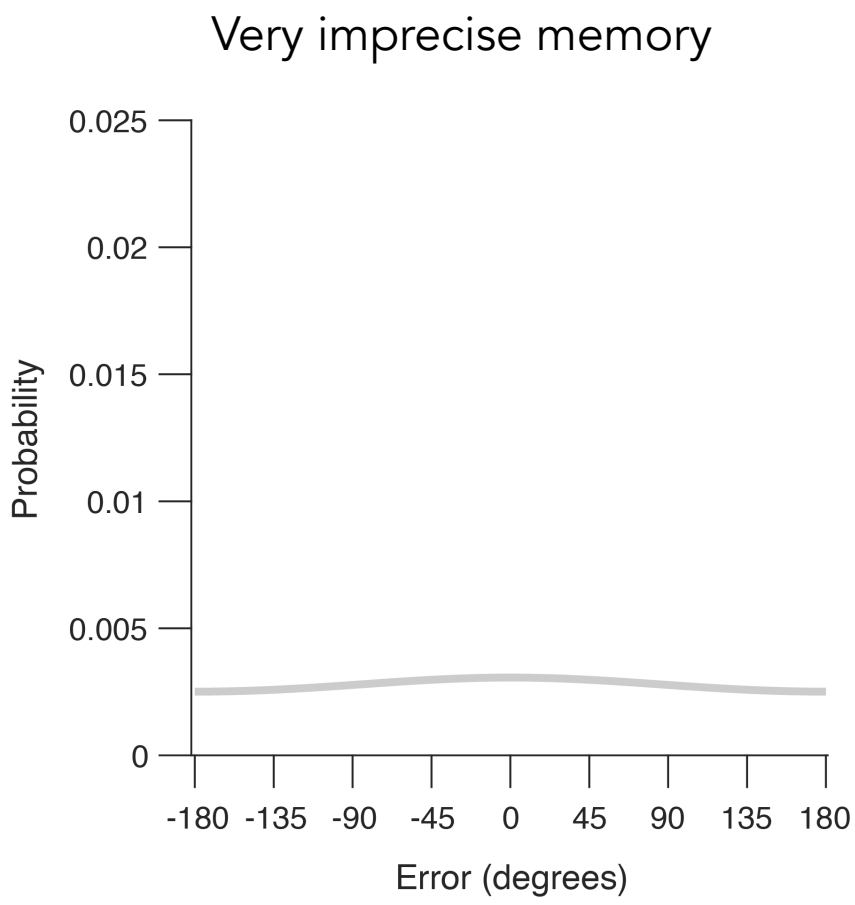
- Variable precision models (van den Berg et al., 2012)

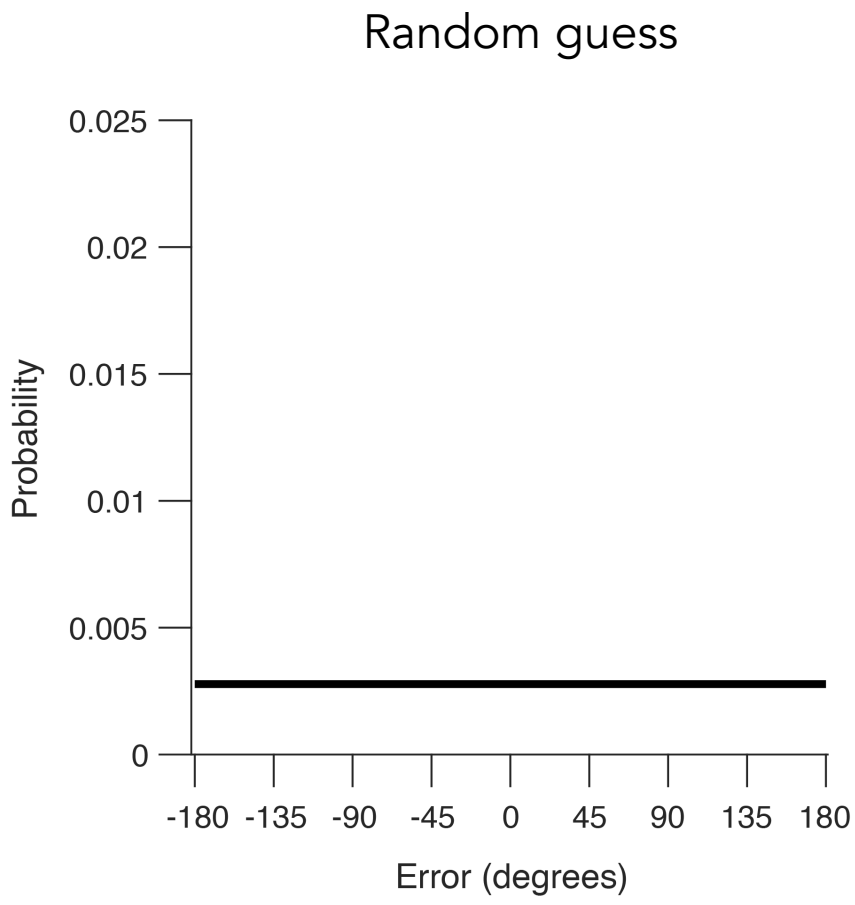


Van den Berg, R., Shin, H., Chou, W. C., George, R., & Ma, W. J. (2012). Variability in encoding precision accounts for visual short-term memory limitations. *Proceedings of the National Academy of Sciences*, *109*(22), 8780-8785.

The issue

- A very imprecise memory response can mimic a random guess





Whole-report recall task

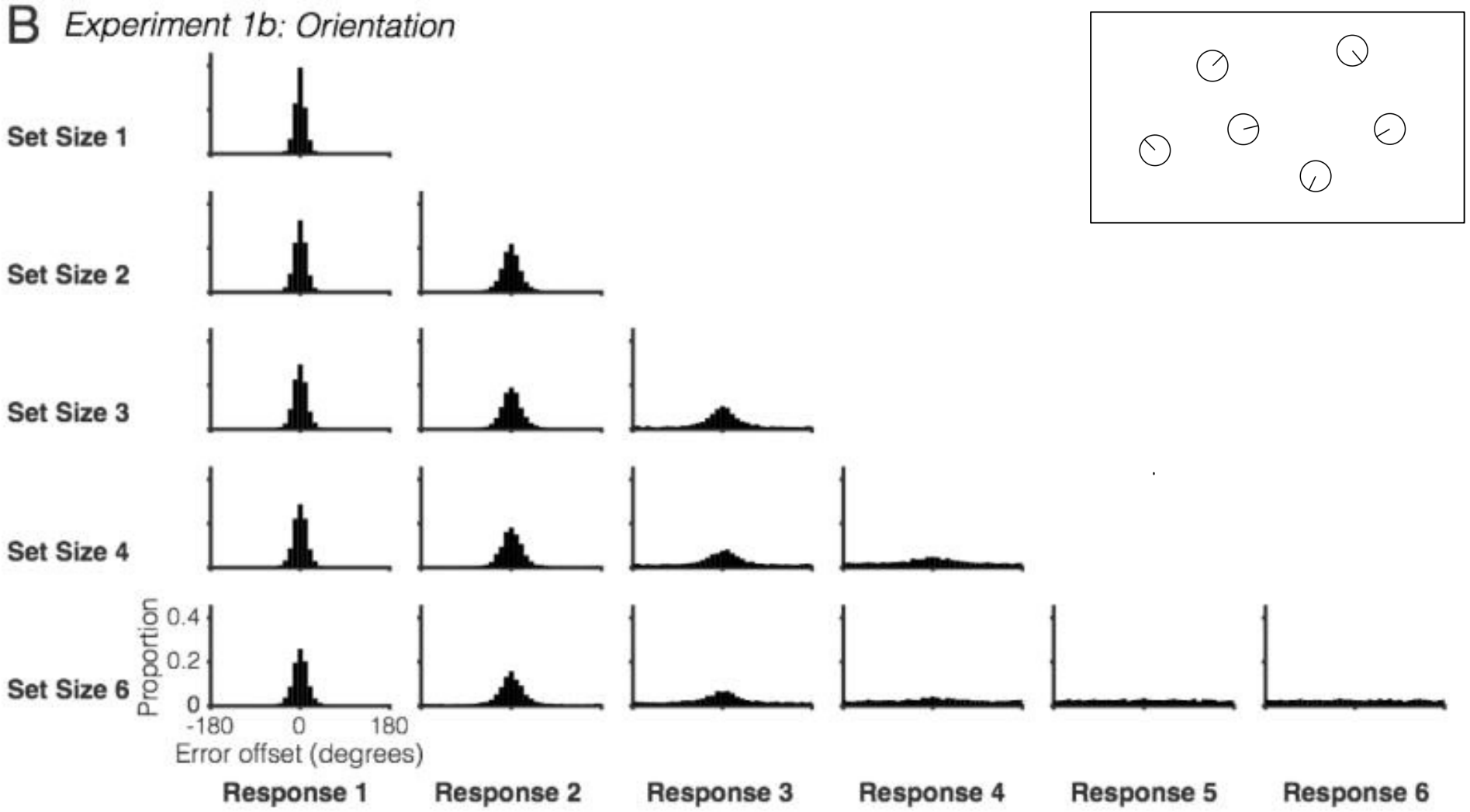
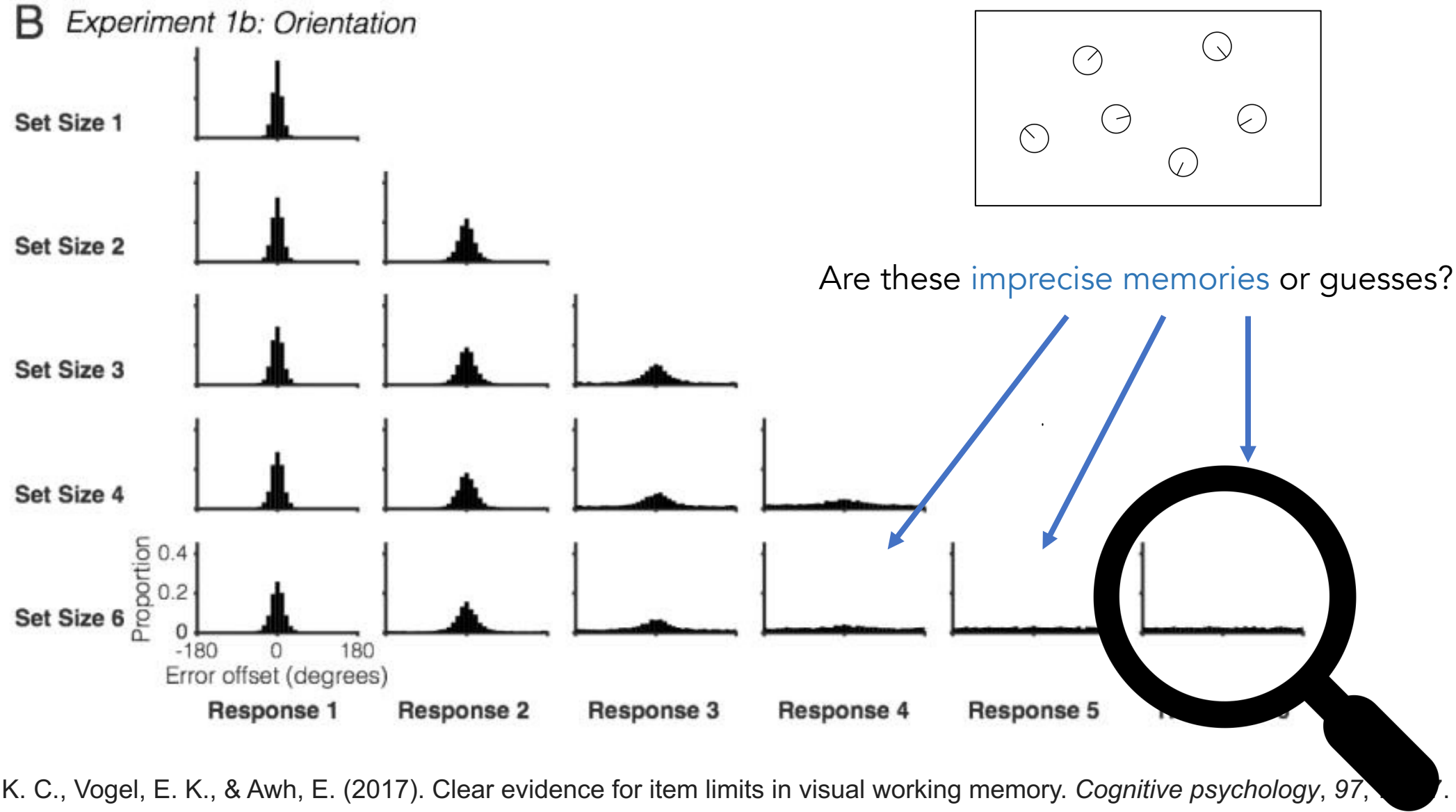


Figure from Adam, K. C., Vogel, E. K., & Awh, E. (2017). Clear evidence for item limits in visual working memory. *Cognitive psychology*, *97*, 79-97.

The issue



Are these imprecise memories or guesses?

Figure from Adam, K. C., Vogel, E. K., & Awh, E. (2017). Clear evidence for item limits in visual working memory. *Cognitive psychology*, *97*, [illegible].

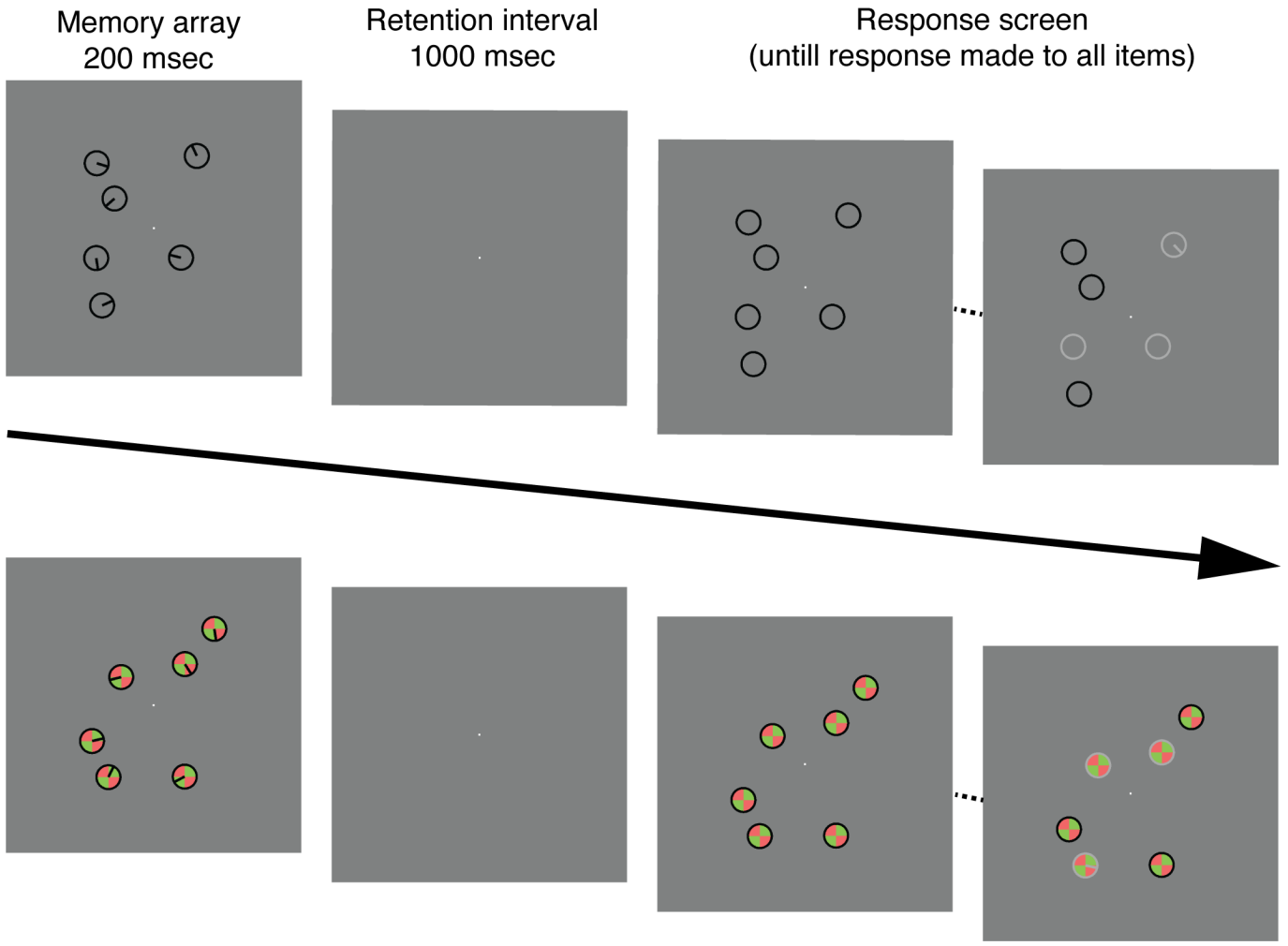
Our solution

- A supposed fundamental difference between these models is the existence of guessing

- Create an experimental paradigm where guesses are clearly distinct from imprecise memories
  - Have guesses produce a different distribution to a uniform distribution

Experiment design

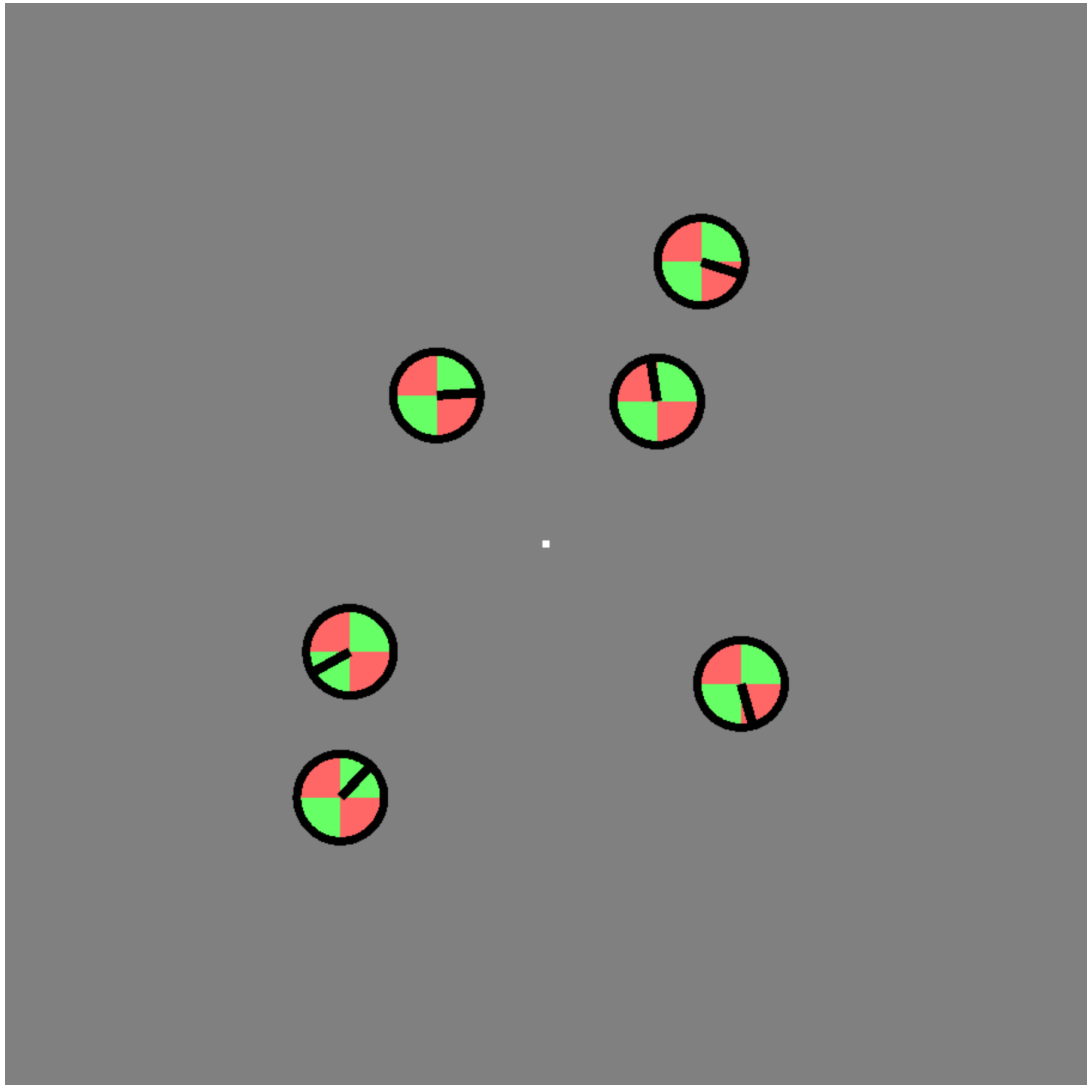
- Whole-report of six orientations
- Experiment 1 (n = 40)
  - 120 trials with colored quadrant backgrounds
  - 80 trials with no background
- Experiment 2 (n = 30)
  - 160 trials with the colored quadrant background rotated 45 degrees



What will guesses look like?

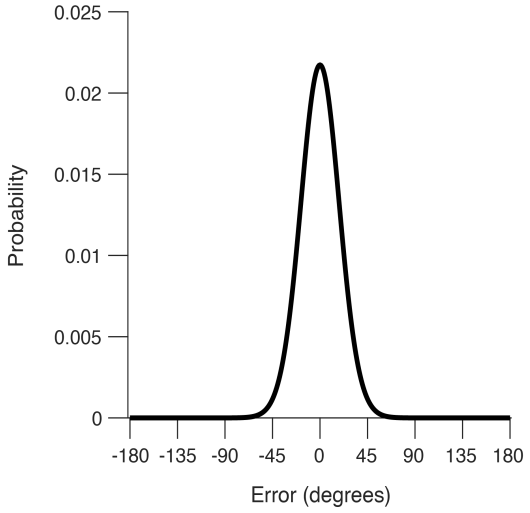
- We expect participants to respond towards the middle of the colored quadrants

- A response that is independent to the presented angle

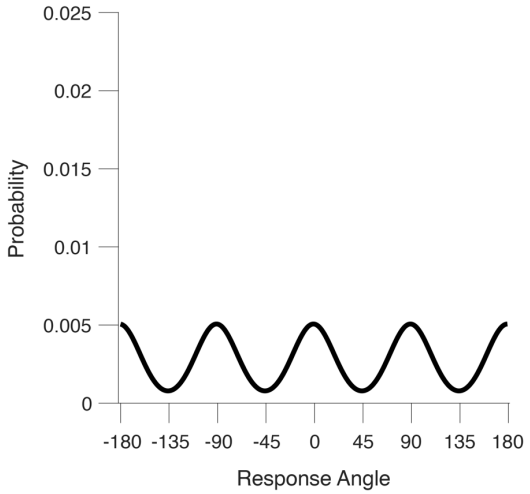


What we predict we will observe

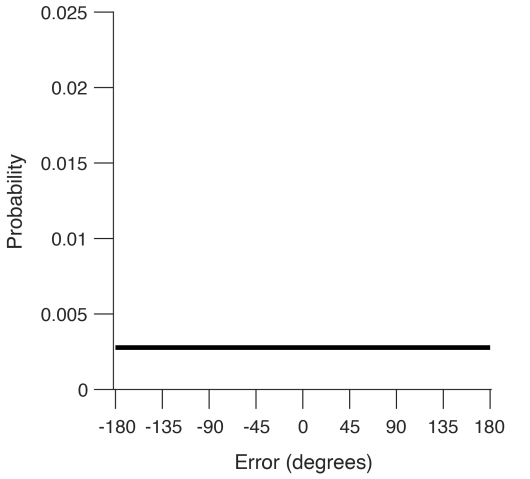
Memory response
(Von Mises)

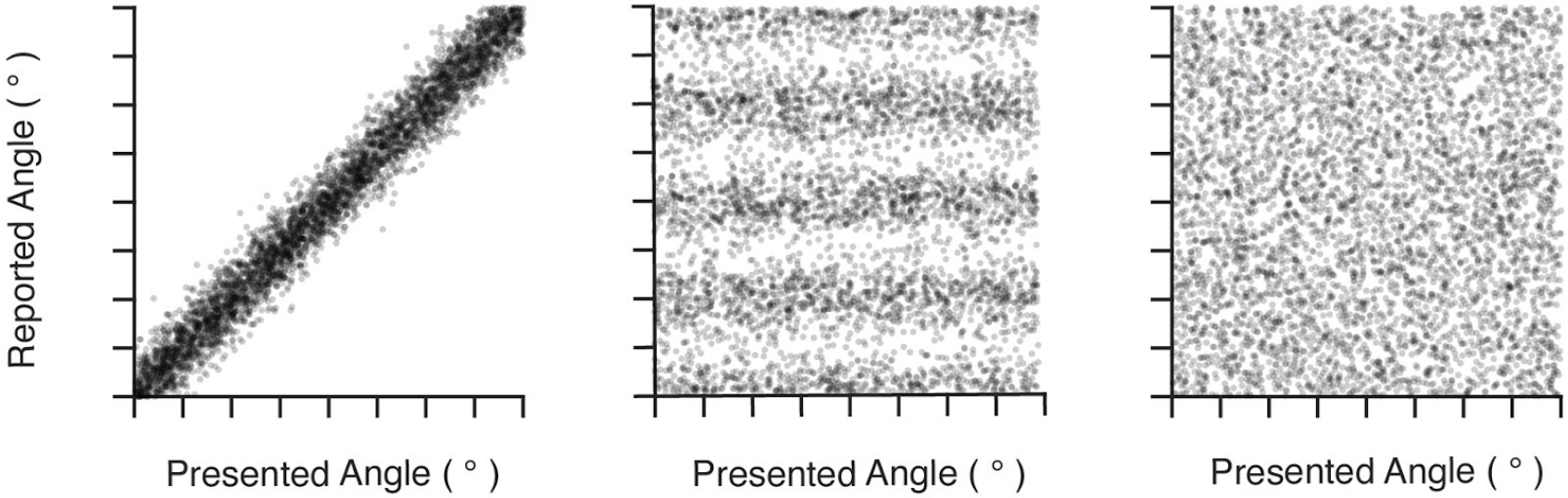


Guess response
(Guess Bands)

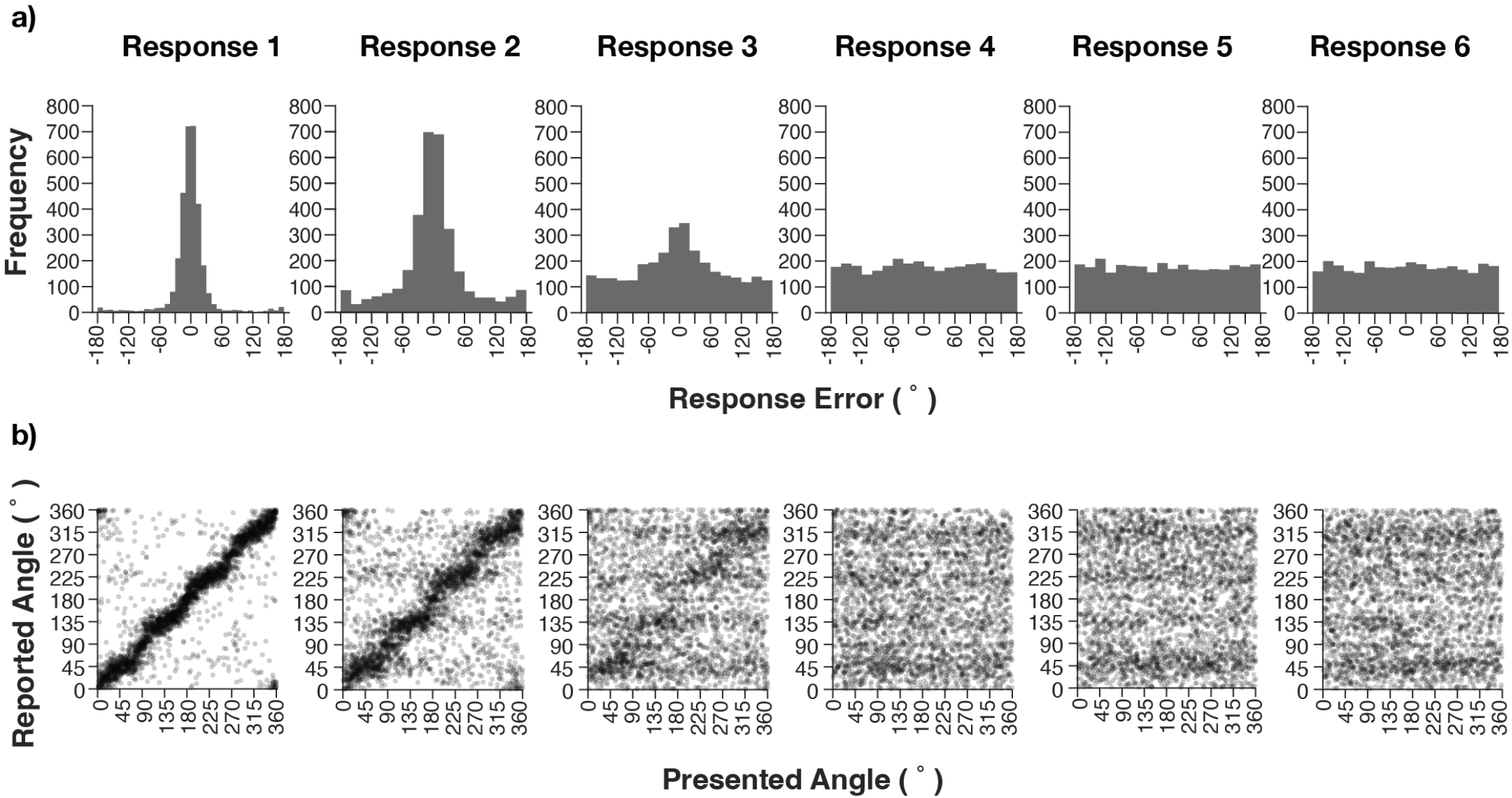


Random response
(Uniform)

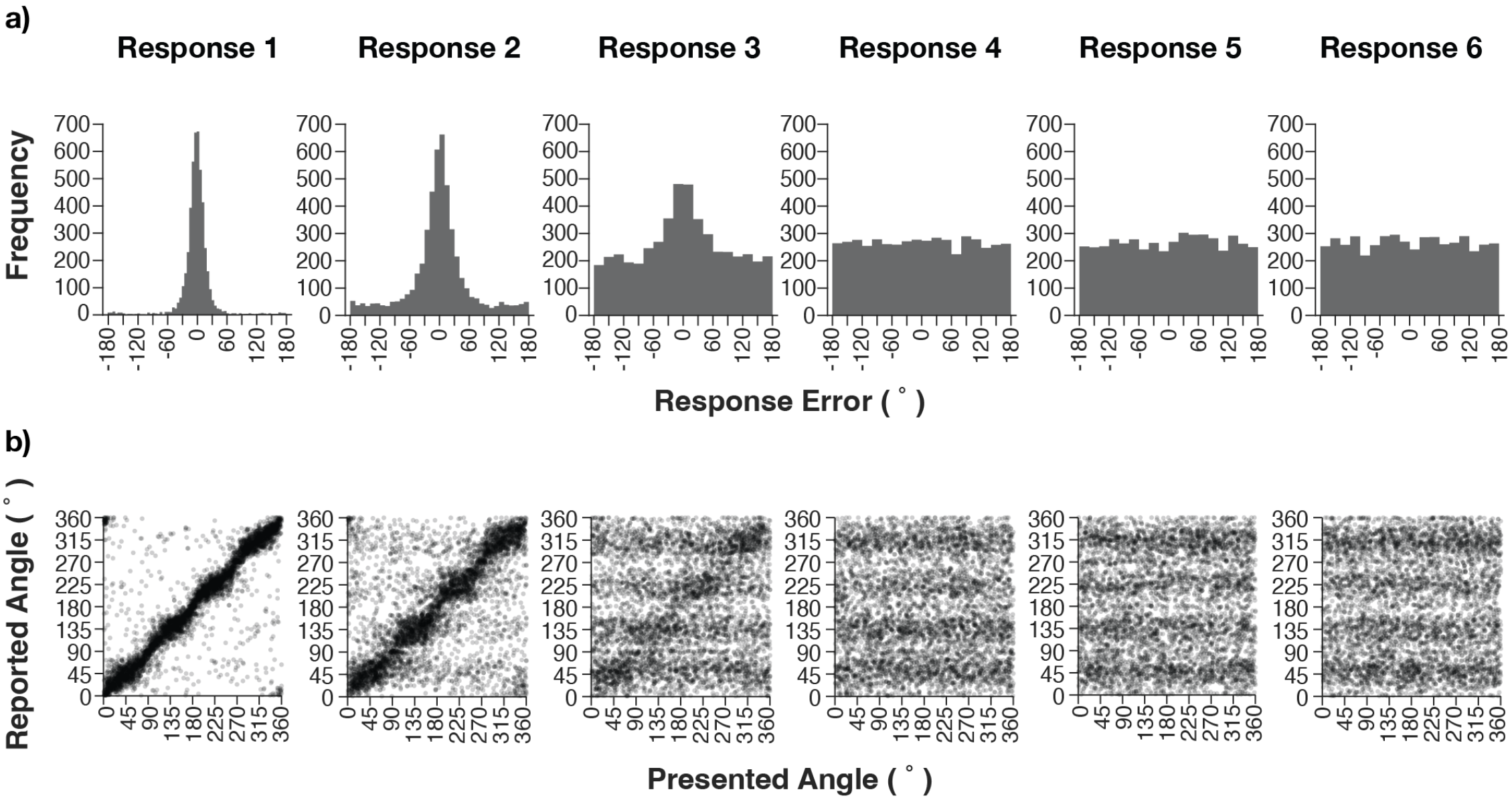




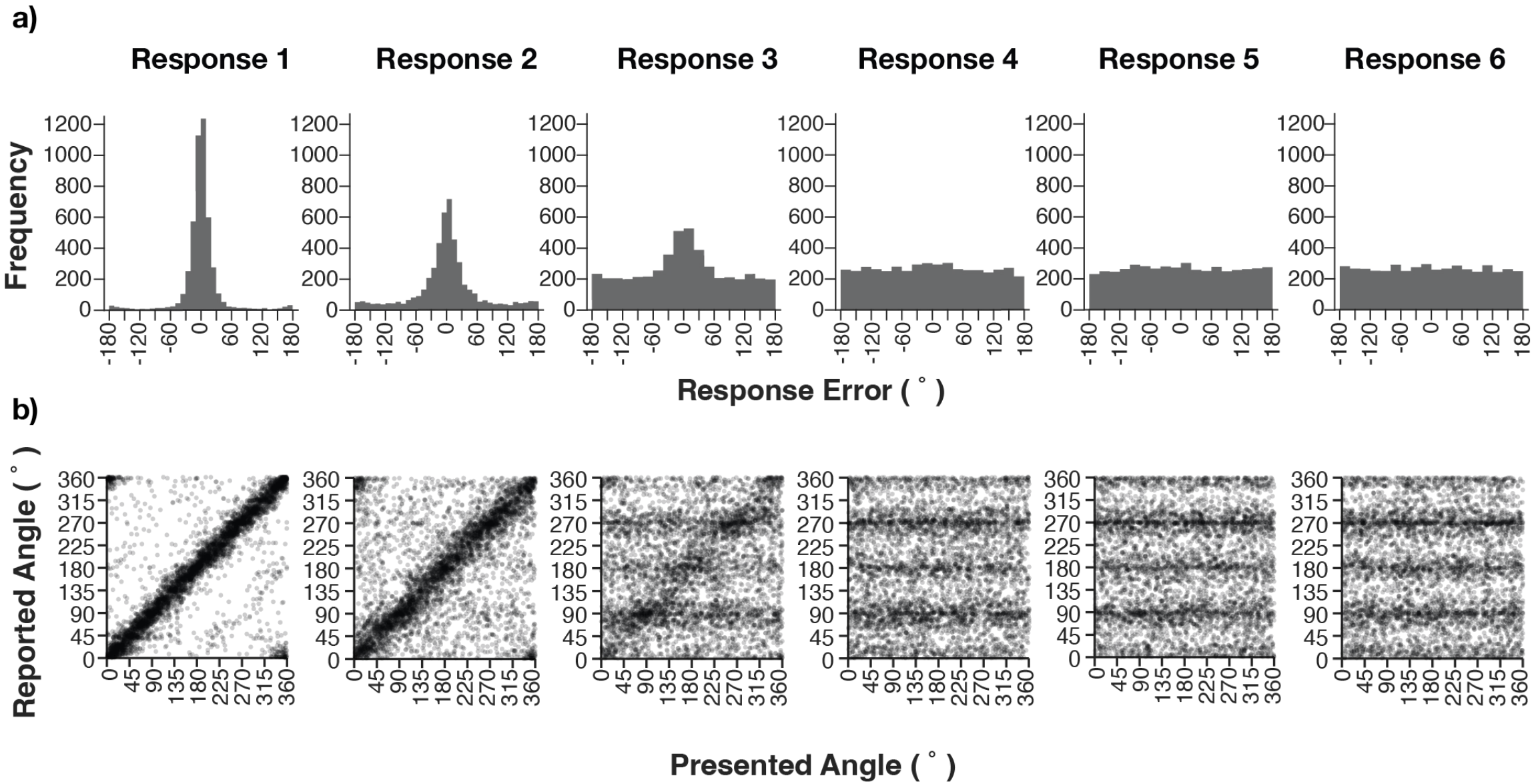
Experiment 1 Results – Standard condition



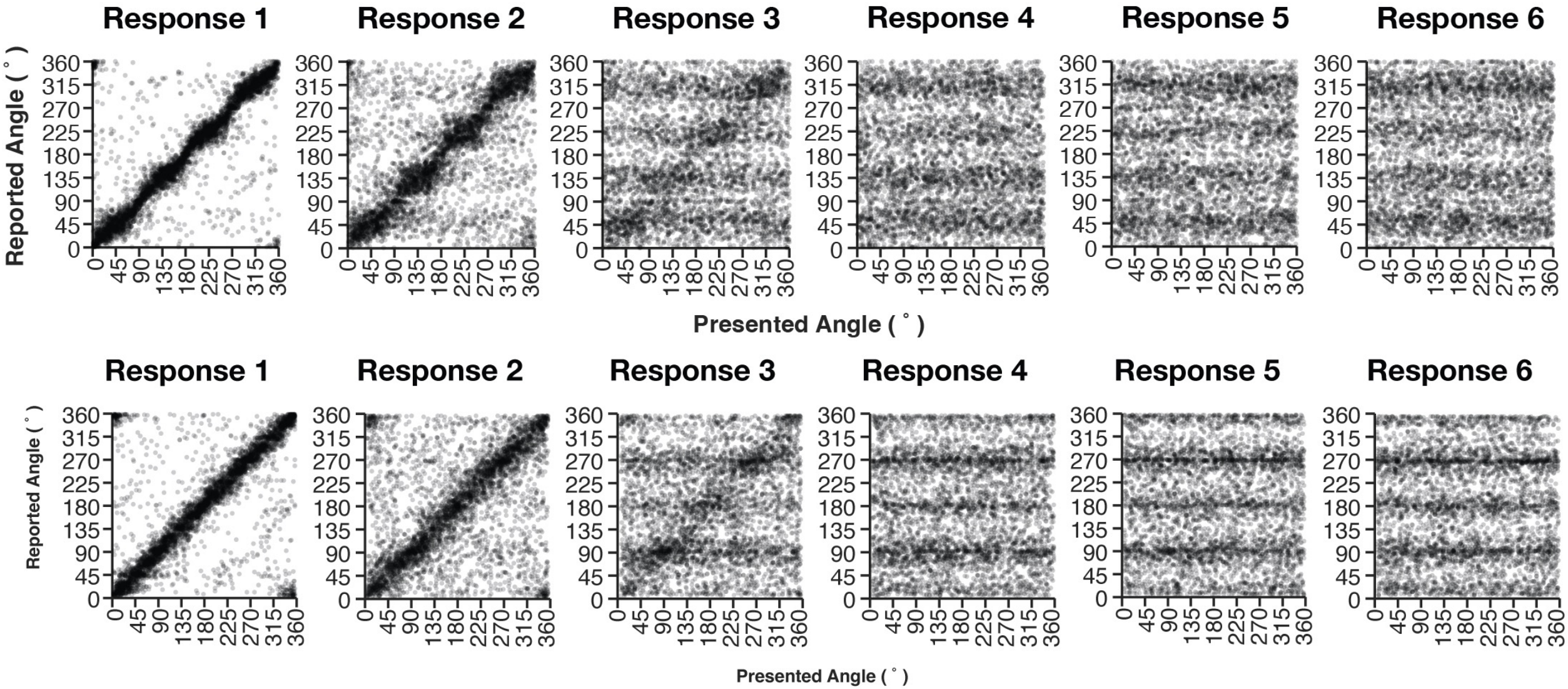
Experiment 1 Results – Background condition



Experiment 2 Results

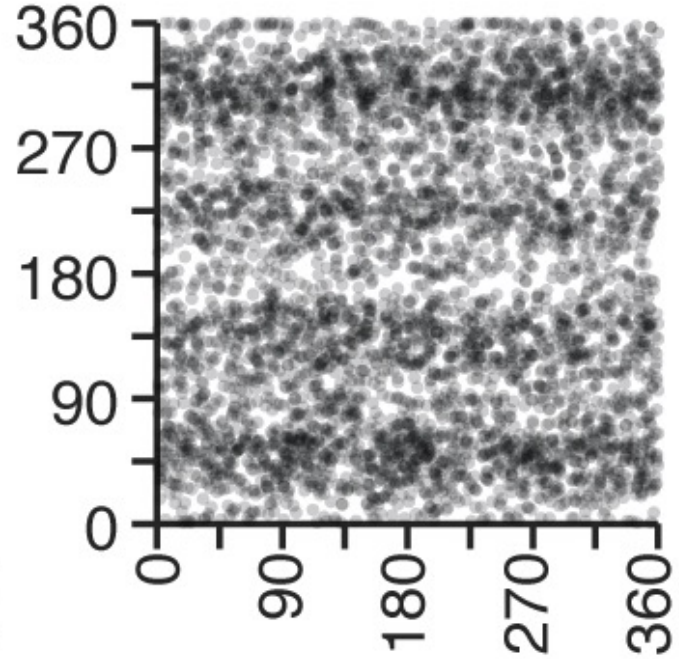


Clear visual evidence for ‘guess bands’



Conclusions

- We found evidence for guesses that cannot straightforwardly described as an imprecise memory
  - In line with an item-based capacity limit
- But the pattern of results can be explained by a resource model
  - One that includes an *ad hoc* change to incorporate priors
  - There may still be a "working memory" masked by the guess responses



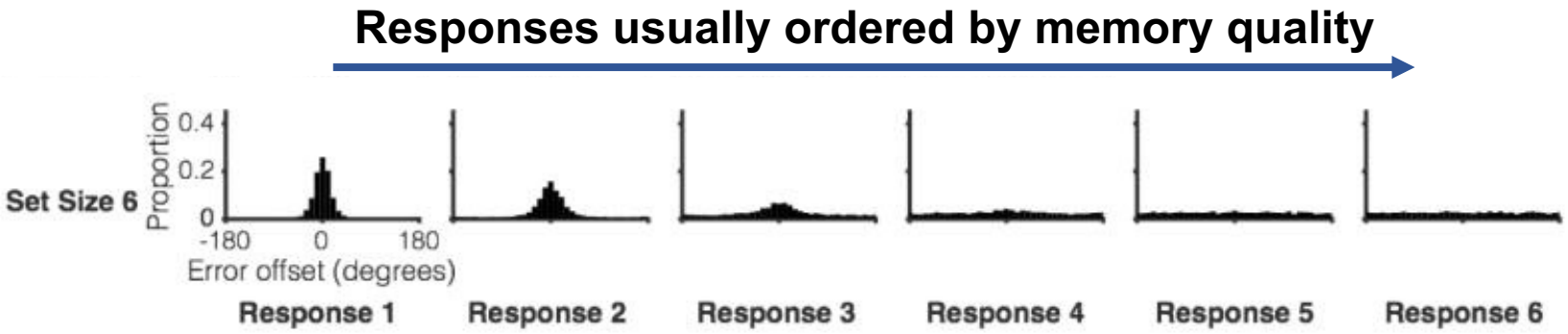
Conjunction whole-report

Krystian Loetscher

Ed Awh

Introducing the conjunction whole-report paradigm

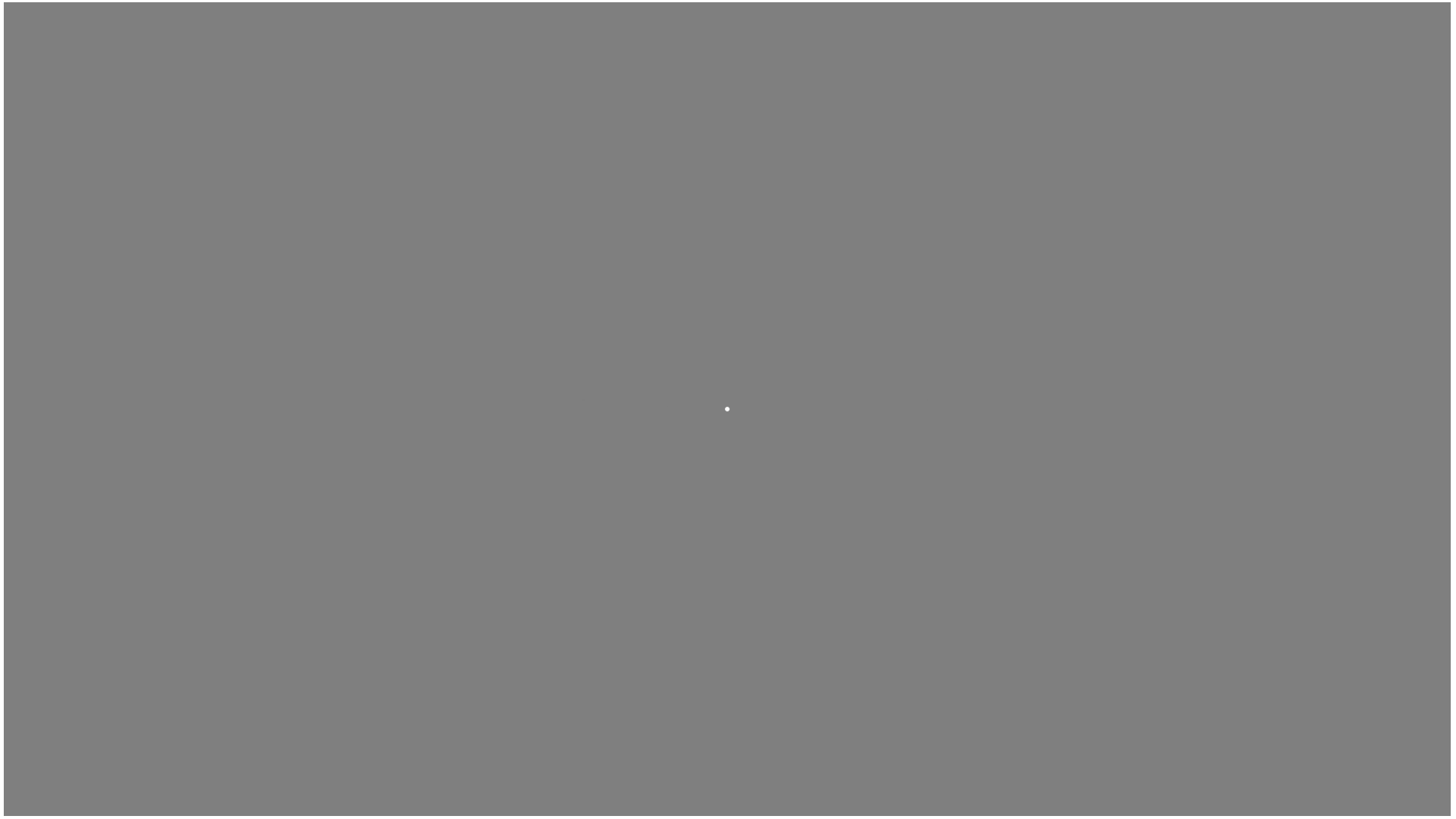
- Test recall for all items rather than just the one item (Adam et al., 2017)



- The first whole-report experiments with conjunction stimuli
- Response interface that collects both features with one click (Sone et al., 2021)

Figure from Adam, K. C. S. et al. (2017) https://doi.org/10.1016/j.cogpsych.2017.07.001
Sone, H. et al. (2021) https://doi.org/10.1016/j.cognition.2020.104579

Conjunction whole-report



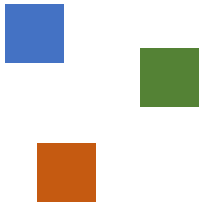
Sone, H., Kang, M. S., Li, A. Y., Tsubomi, H., & Fukuda, K. (2021). Simultaneous estimation procedure reveals the object-based, but not space-based, dependence of visual working memory representations. *Cognition*, *209*, 104579.

What *is* visual working memory?

Object-based theory

"slot models"

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



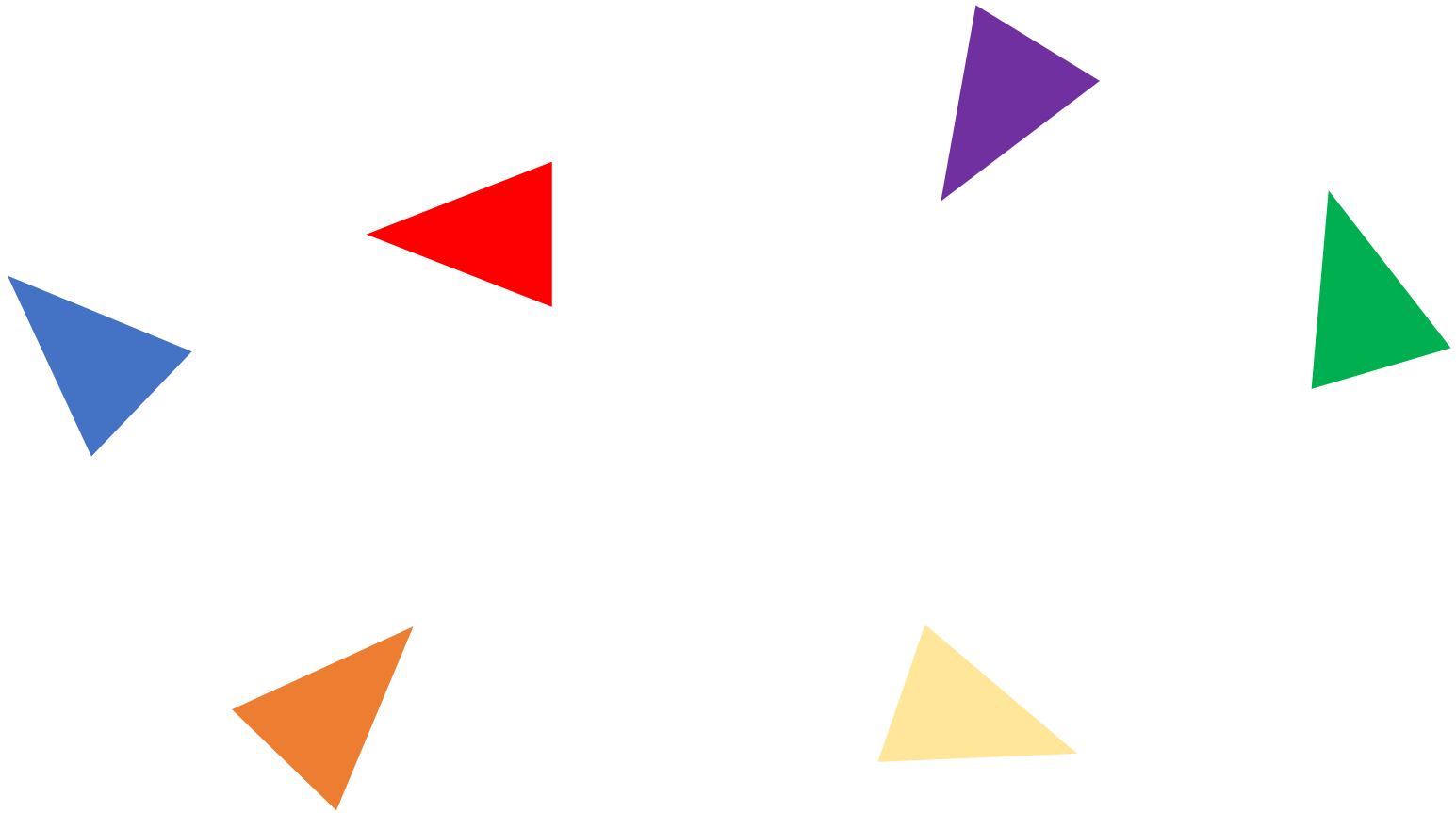
Feature-based theory

"resource models"

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

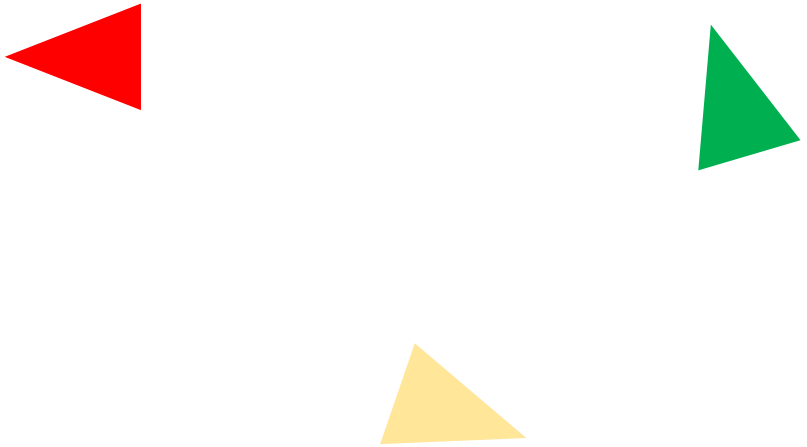
Luck, S. J., & Vogel, E. K. (1997). https://doi.org/10.1038/36846
Zhang, W., & Luck, S. J. (2008). https://doi.org/10.1038/nature06860
Alvarez, G. A., & Cavanagh, P. (2004). https://doi.org/10.1111/j.0963-7214.2004.01502006.x
Wilken, P., & Ma, W. J. (2004). https://doi.org/10.1167/4.12.11

What is the unit of working memory?



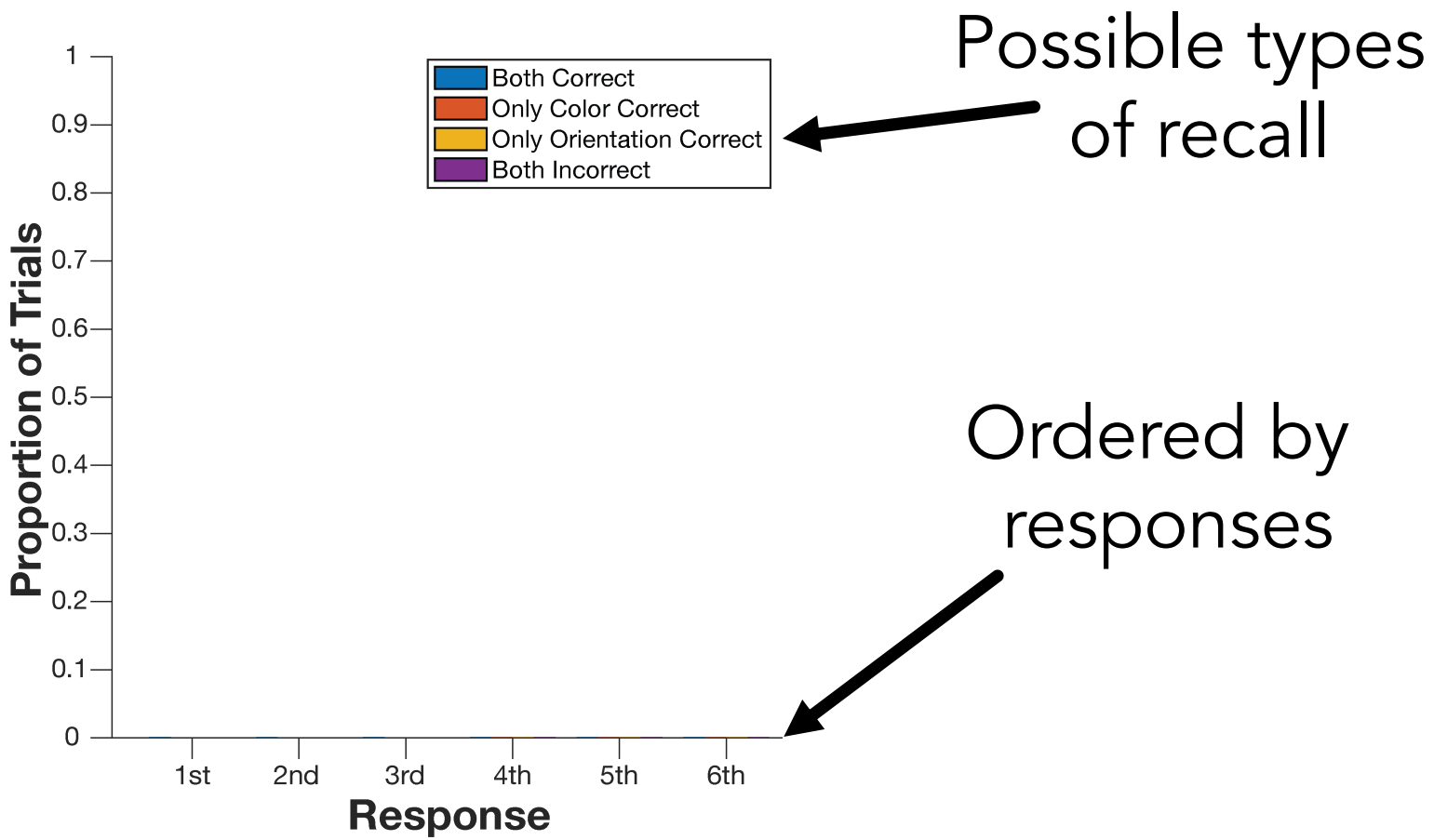
A *specific* object-based model – strong objects

- Fixed object capacity limit
- Lossless representations (“all-or-none”)
- No impact of complexity (additional features)



Has anyone ever truly believed this? Anyhow, an early rejection of this model: Olson, I. R. and Jiang, Y. (2002) https://doi.org/10.3758/BF03194756

Conjunction whole-report

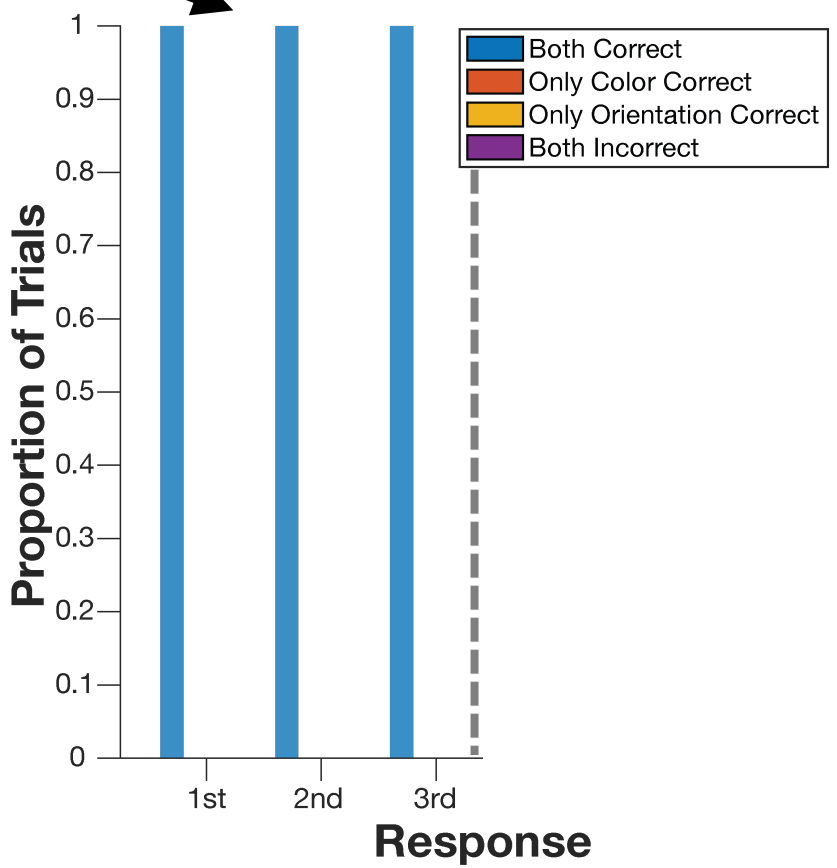


Possible types of recall

Ordered by responses

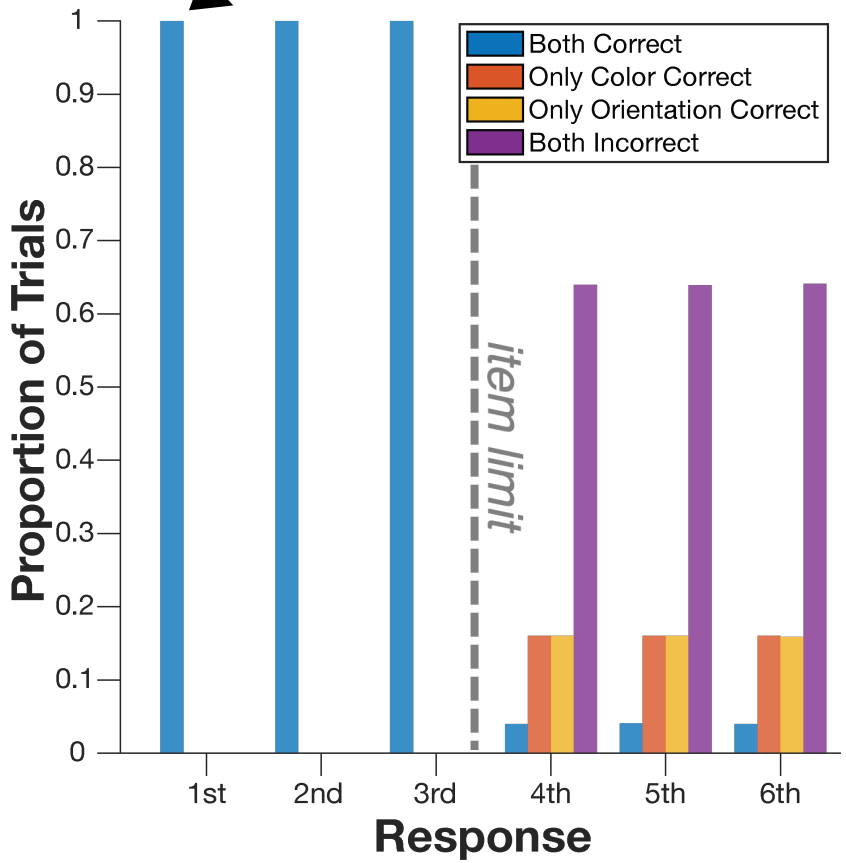
A *specific* slot model – strong objects

Perfect recall within item limit



A *specific* slot model – strong objects

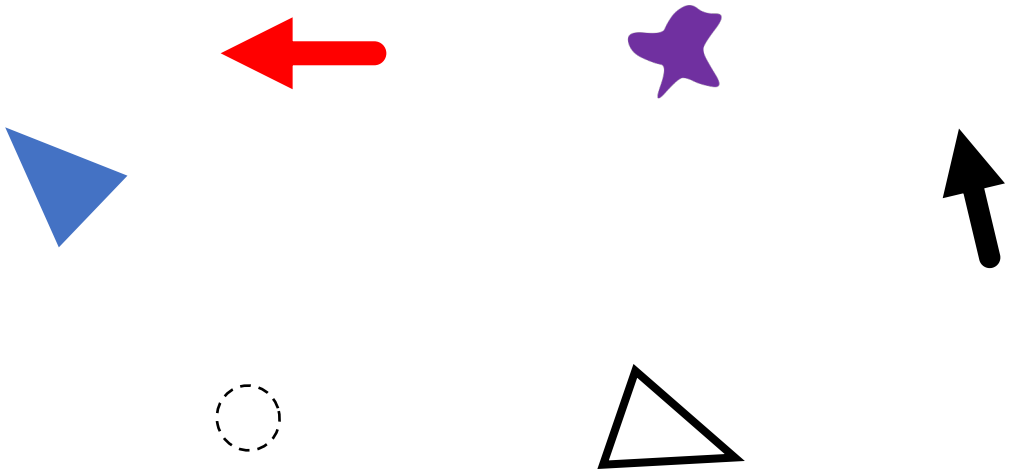
Perfect recall within item limit



Guessing for remaining responses

A *specific* resource model – independent features

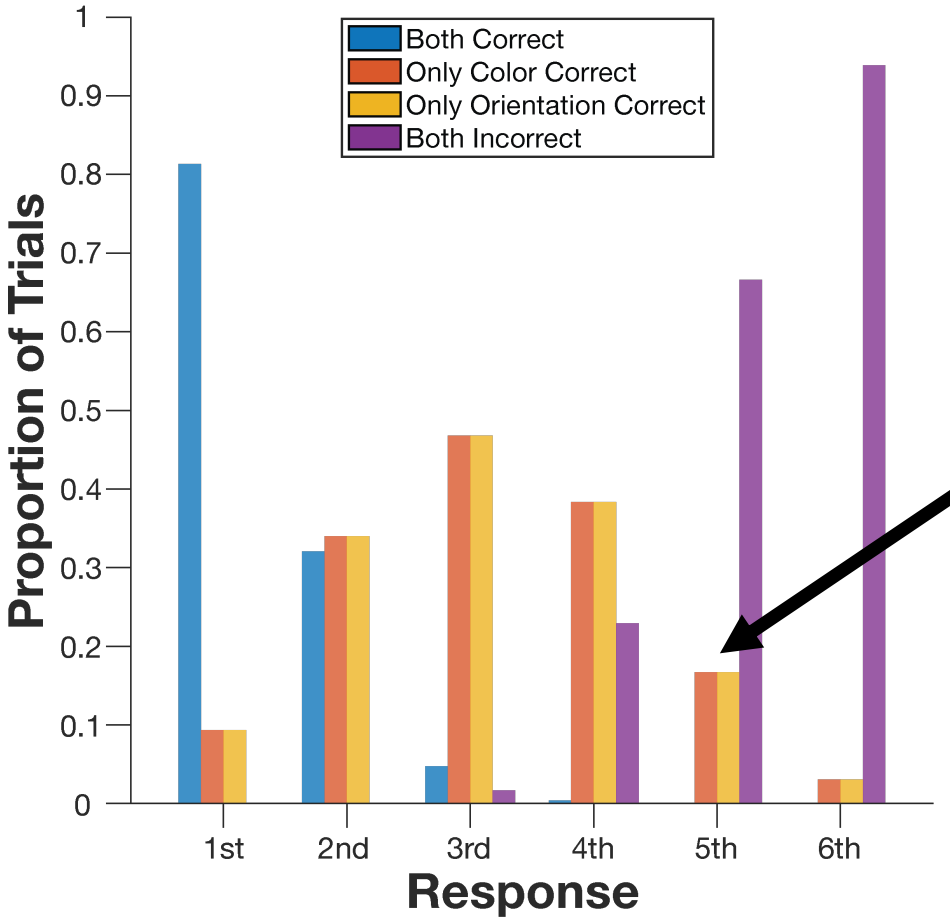
- Working memory resources are distributed to all items in the array
- Feature storage is not constrained by which objects contain the features
  - Probability of successful feature storage is independent of objecthood



Bundesen, C. (1990) https://doi.org/10.1037/0033-295X.97.4.523

A *specific* resource model – independent features

Feature storage independent of objecthood



Accurate recall distributed across *all* responses

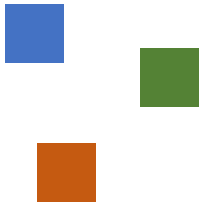
What *is* visual working memory?

- An enduring theoretical framework has been

Object-based theory

"slot models"

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



versus

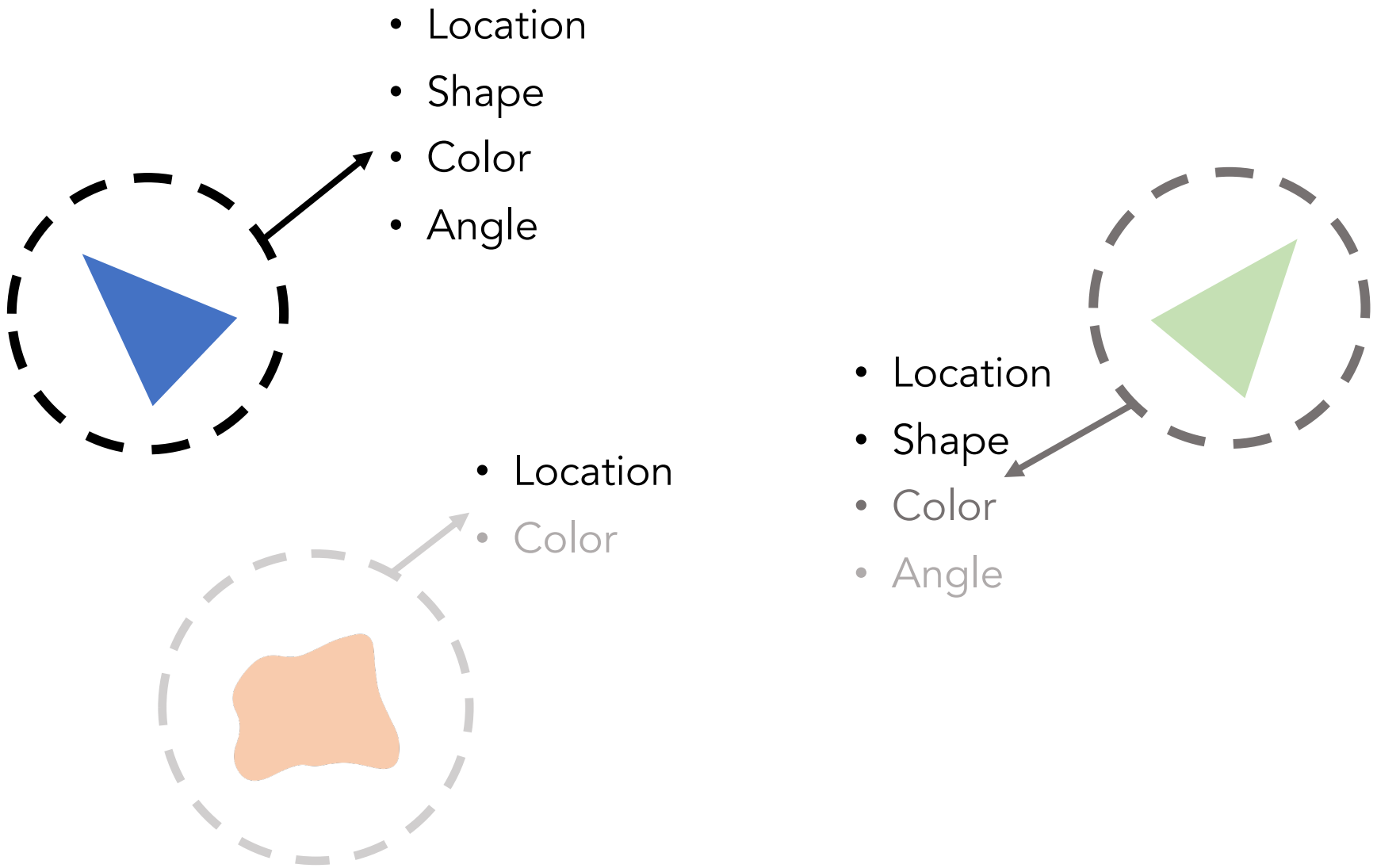
Feature-based theory

"resource models"

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

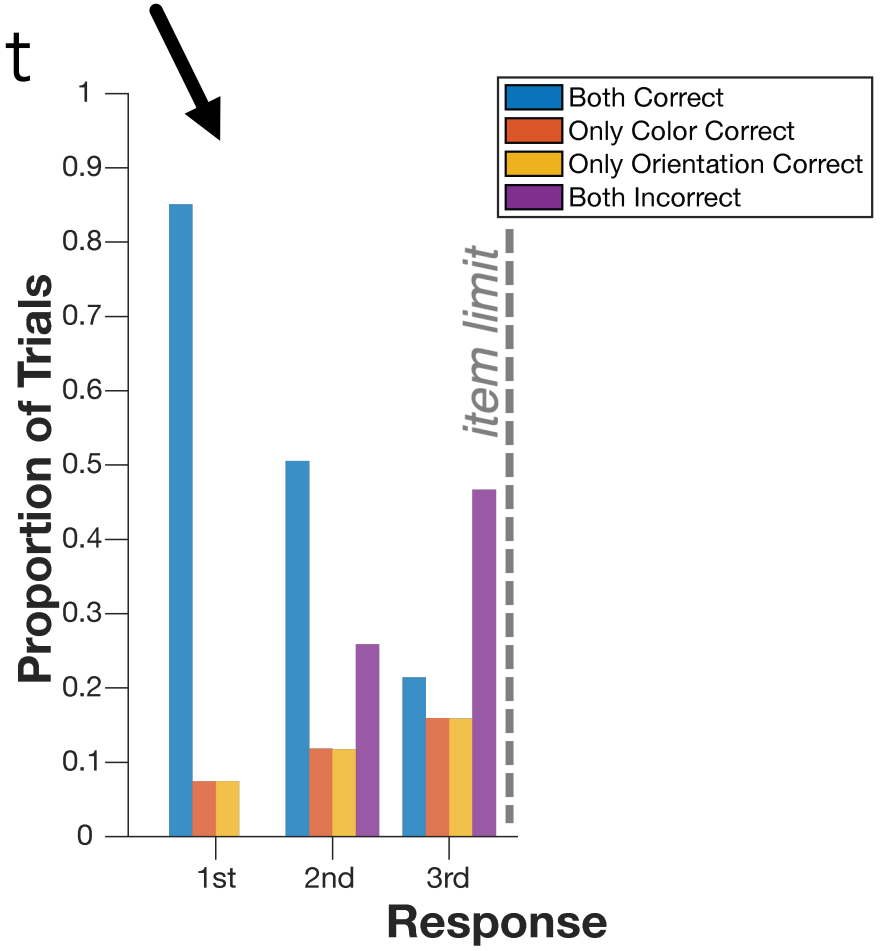
Luck, S. J., & Vogel, E. K. (1997). https://doi.org/10.1038/36846
Zhang, W., & Luck, S. J. (2008). https://doi.org/10.1038/nature06860
Alvarez, G. A., & Cavanagh, P. (2004). https://doi.org/10.1111/j.0963-7214.2004.01502006.x
Wilken, P., & Ma, W. J. (2004). https://doi.org/10.1167/4.12.11

A new model characterization – pointers



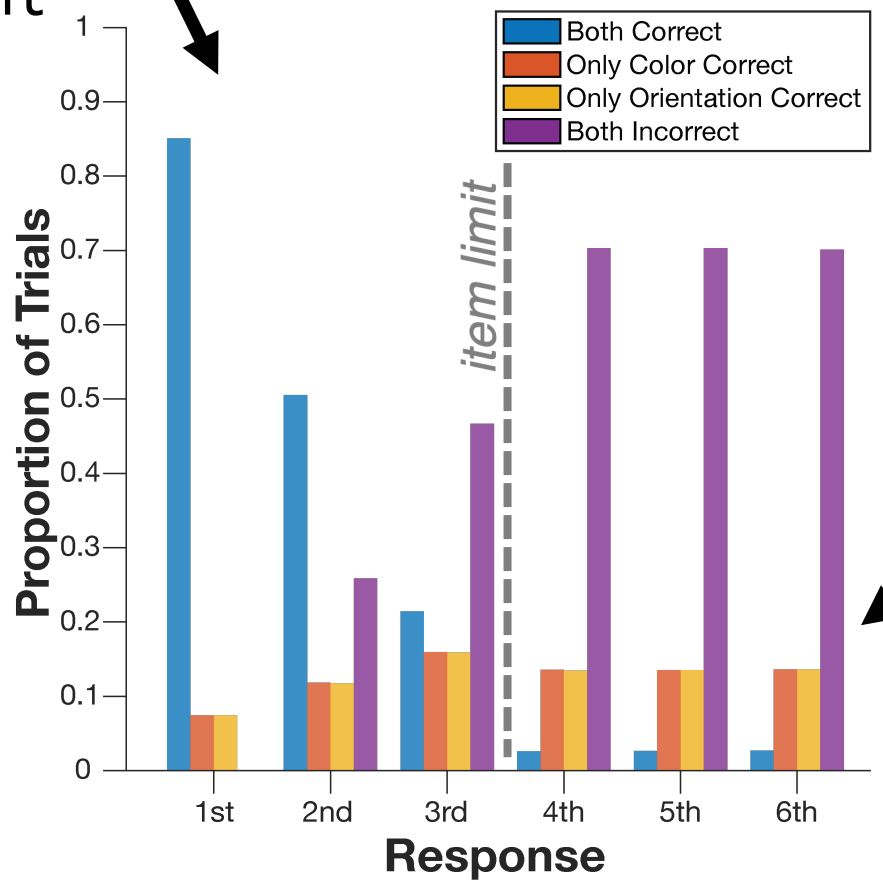
A new model characterization – pointers

Accurate recall constrained within an item limit



A new model characterization – pointers

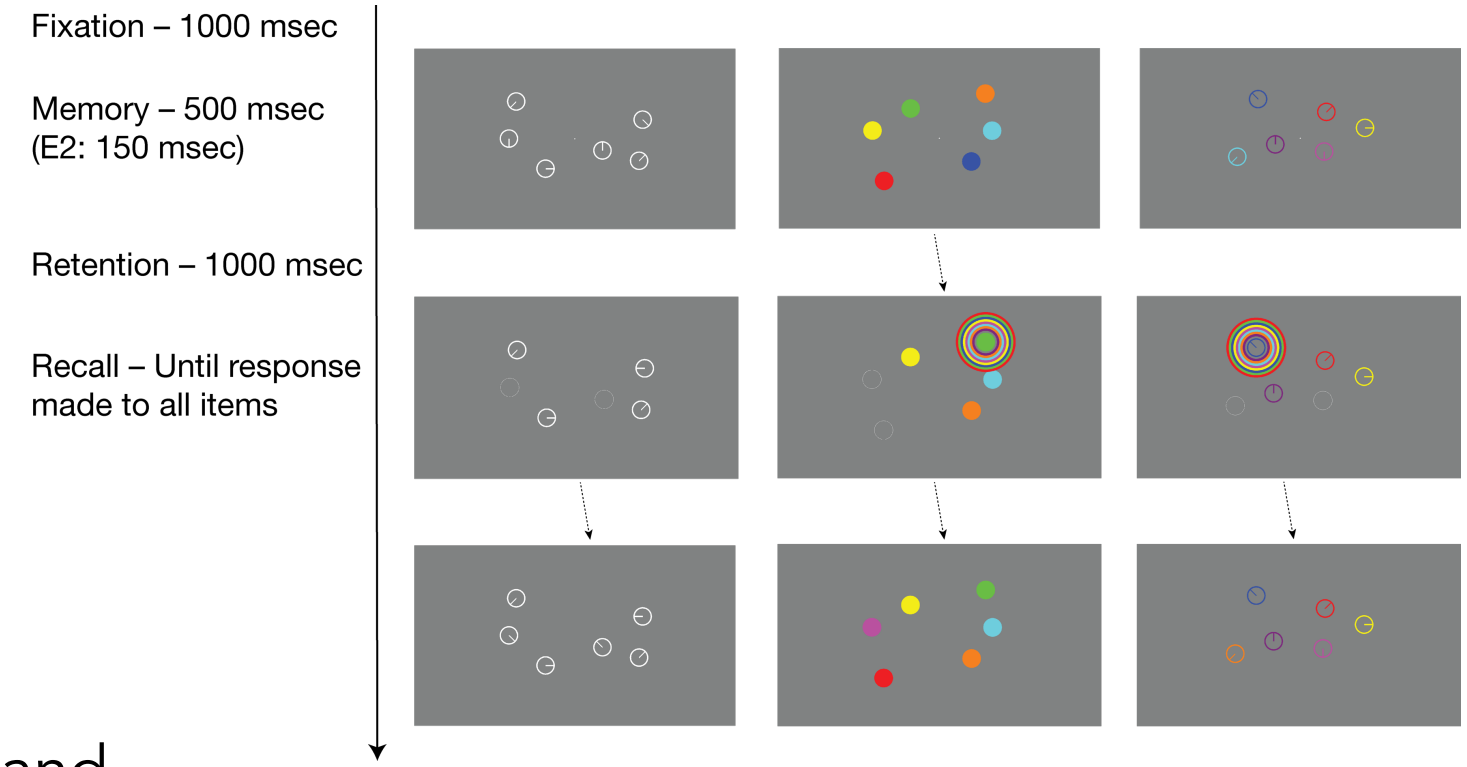
Accurate recall constrained within the item limit



Guessing beyond the item limit

Our conjunction whole-report experiments

- Four experiments (30 subjects each)
  - E1: Colored clock faces
  - E2: Colored clock faces but rapid
  - E3: Colored triangles
  - E4: Colored shapes
- Three conditions (300 trials each)
  - Color only
  - Orientation only or Shape only
  - Conjunction
- Eight discrete colors, orientations, and shapes.



Recall accuracy

Mean Recall	Experiment 1	Experiment 2	Experiment 3	Experiment 4
Colors	3.21 ± 0.74	2.94 ± 0.64		3.61 ± 0.75
Orientations/Shapes	2.79 ± 0.44	2.45 ± 0.45		3.39 ± 0.64
Conjunctions	1.62 ± 0.38	1.38 ± 0.42	1.47 ± 0.44	1.92 ± 0.43

- Memory for conjunction stimuli is not lossless
  - Less conjunctions are fully recalled overall

"It's not objects, it's features!"

Recall accuracy

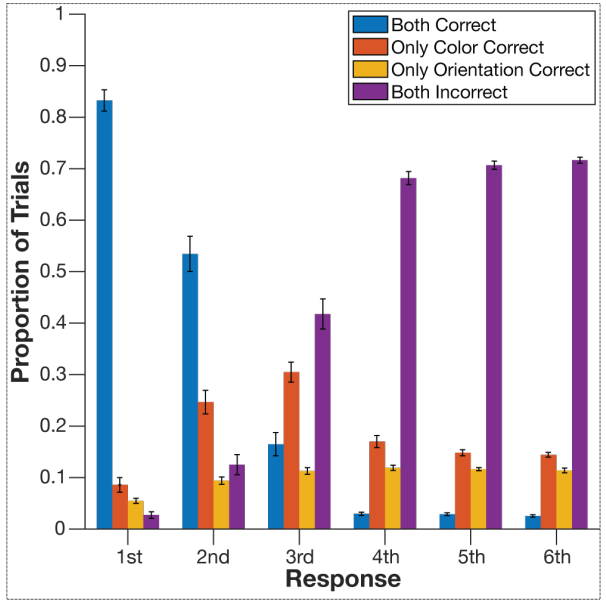
Mean Recall	Experiment 1	Experiment 2	Experiment 3	Experiment 4
Colors	3.21 ± 0.74	2.94 ± 0.64		3.61 ± 0.75
Orientations/Shapes	2.79 ± 0.44	2.45 ± 0.45		3.39 ± 0.64
Conjunctions	1.62 ± 0.38	1.38 ± 0.42	1.47 ± 0.44	1.92 ± 0.43
Features	4.94 ± 0.68	4.52 ± 0.83	5.11 ± 0.65	5.34 ± 0.85

- Memory for conjunction stimuli is not lossless
  - Less conjunctions are fully recalled overall
- But we observe an object-based benefit
  - More features are recalled overall in the conjunction condition compared to the single-feature conditions (~5 features versus ~3 features)

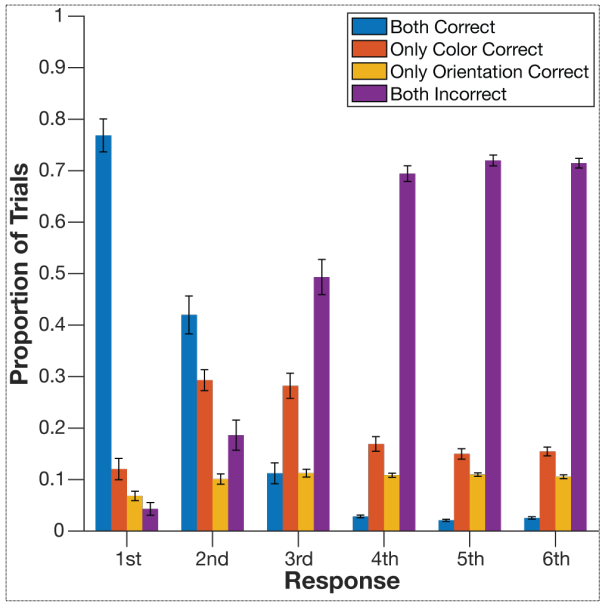
"It's objects, not features!"

Accuracy across responses

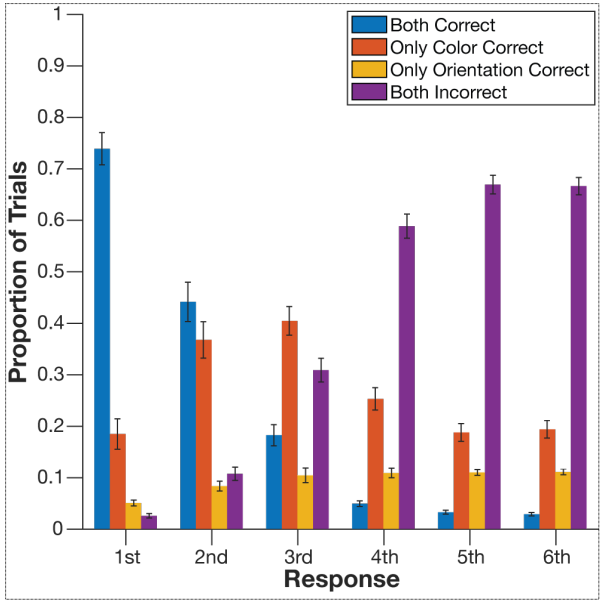
Experiment 1



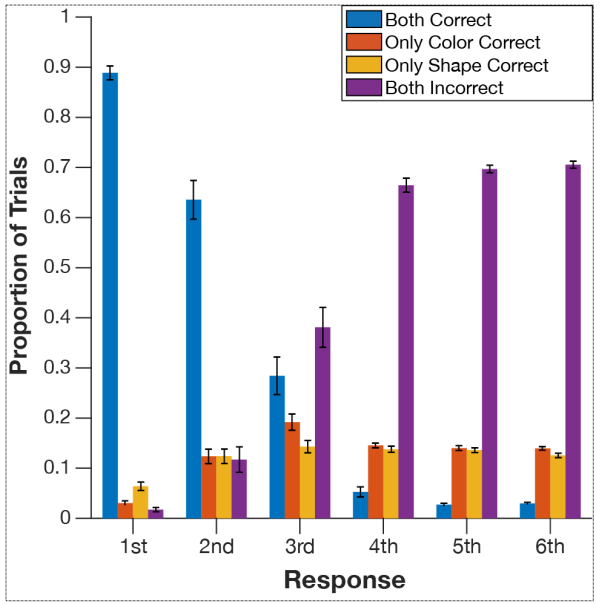
Experiment 2



Experiment 3



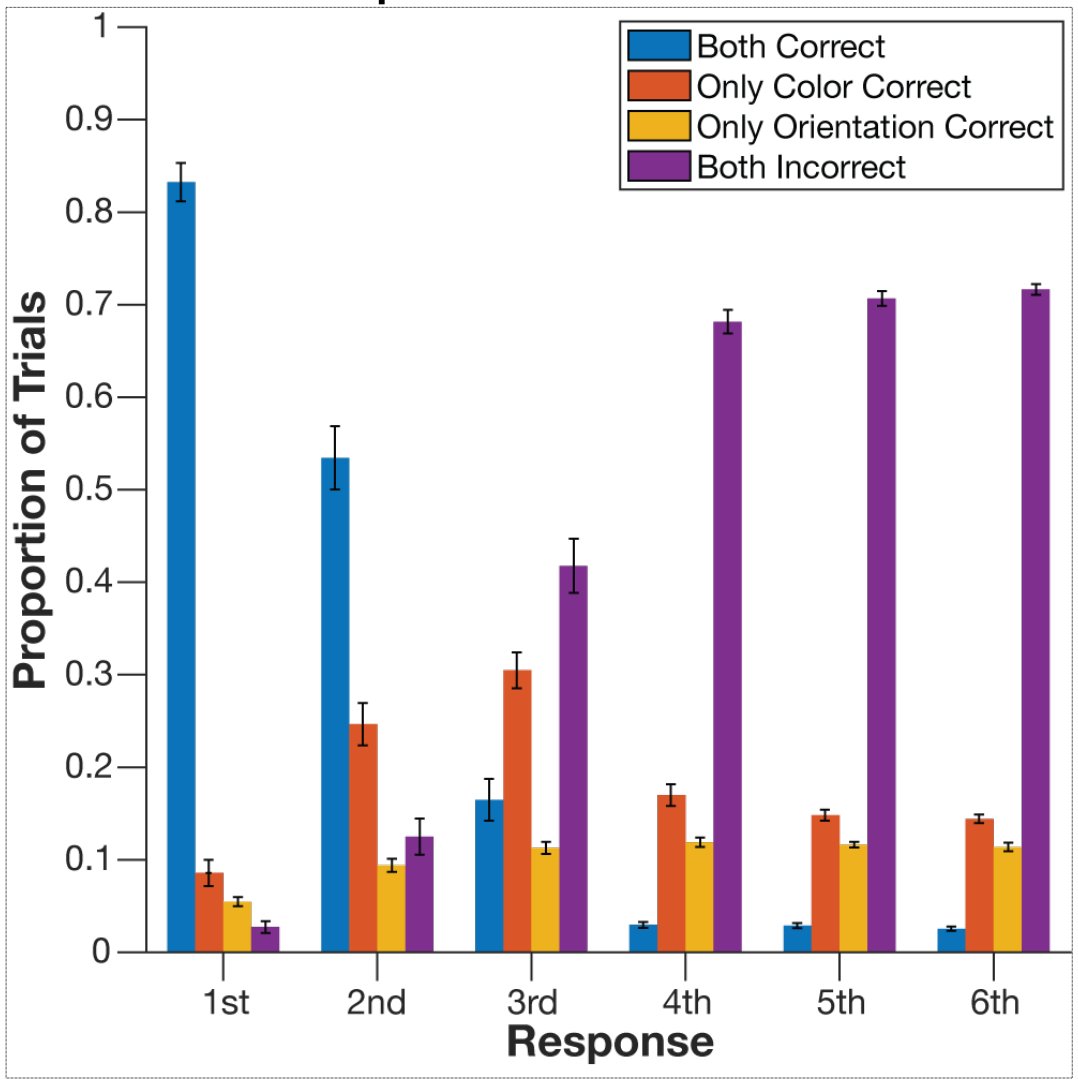
Experiment 4



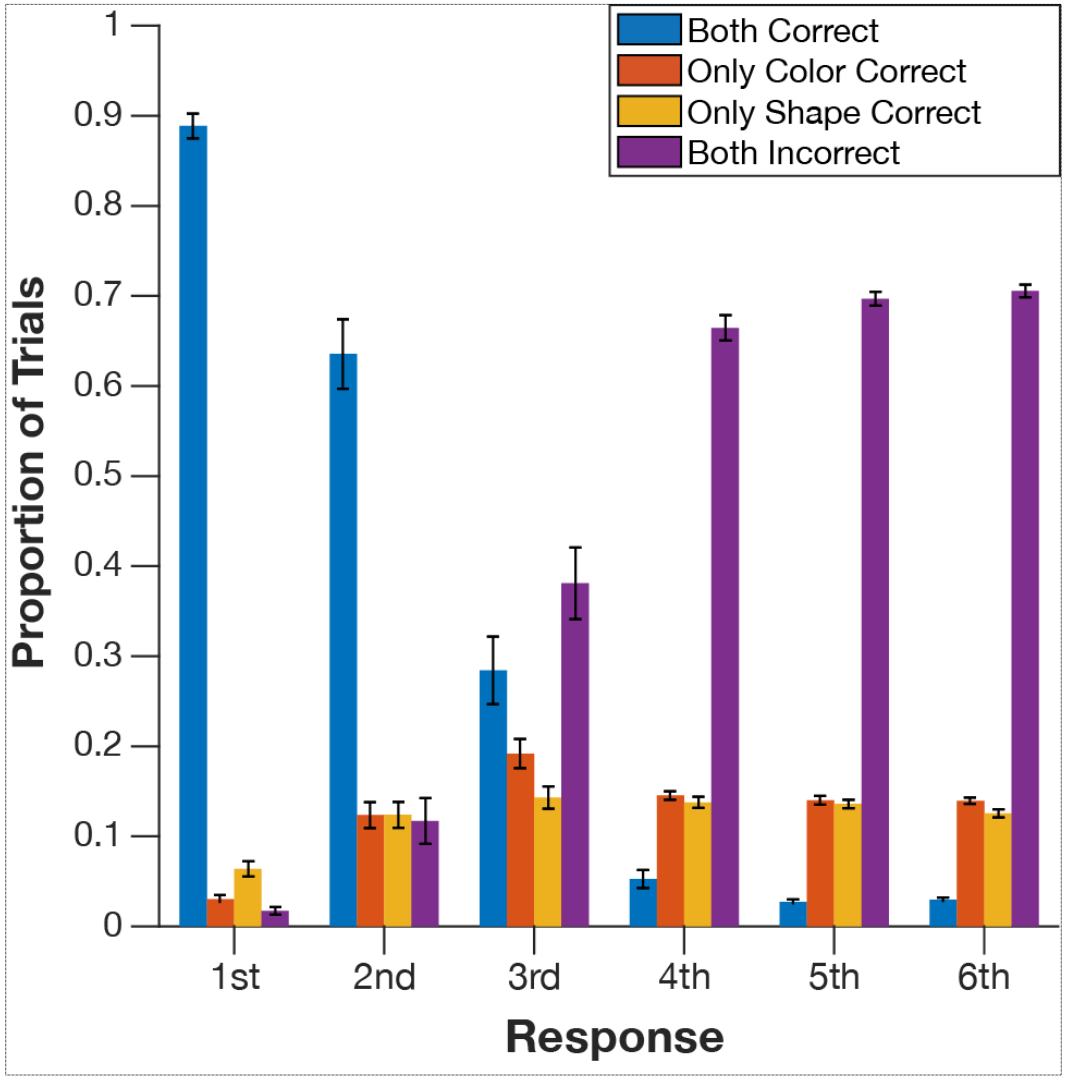
- The same empirical pattern was replicated across four experiments

Accuracy across responses

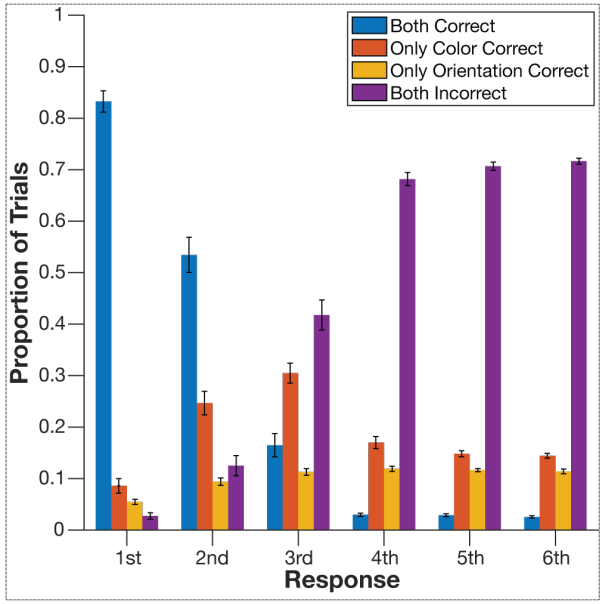
Experiment 1



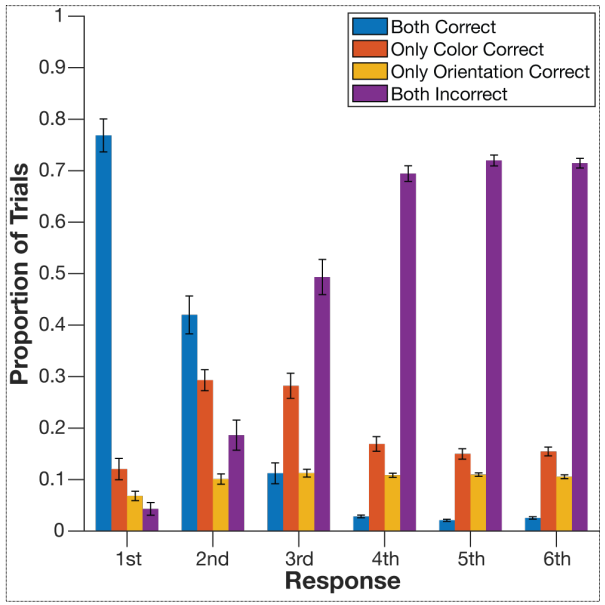
Experiment 4



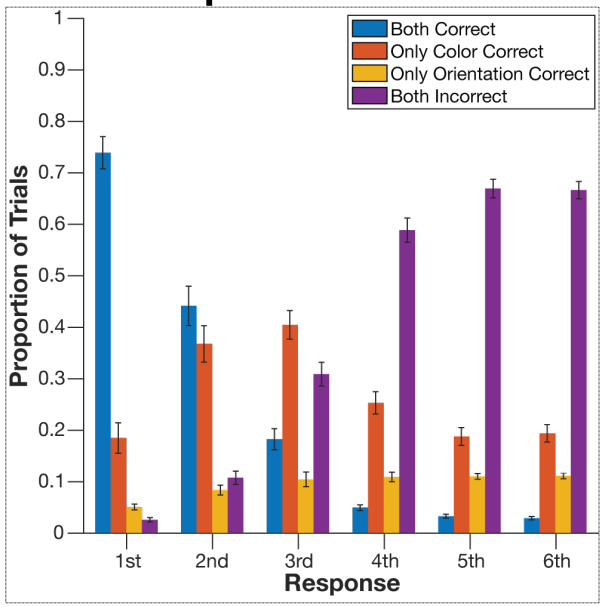
Experiment 1



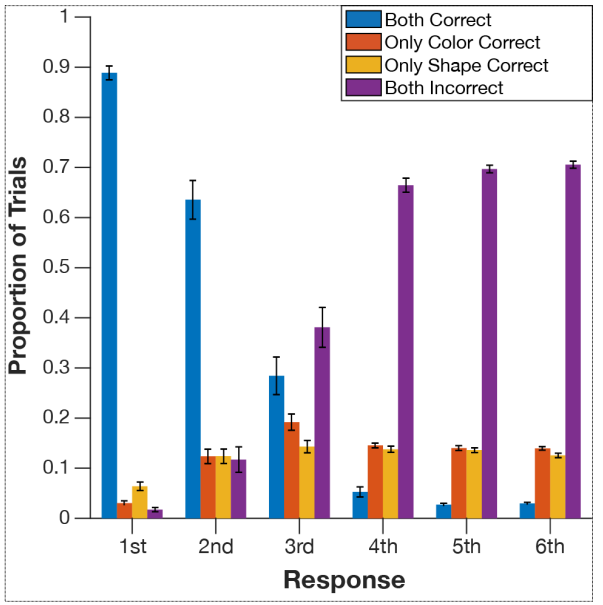
Experiment 2



Experiment 3

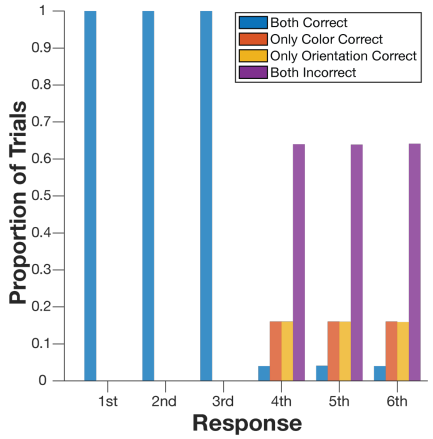


Experiment 4



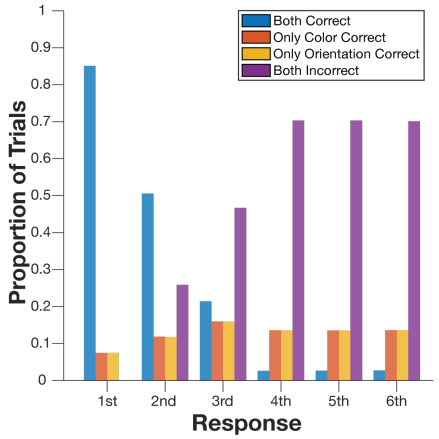
Strong Object Model

Accurate storage of three objects



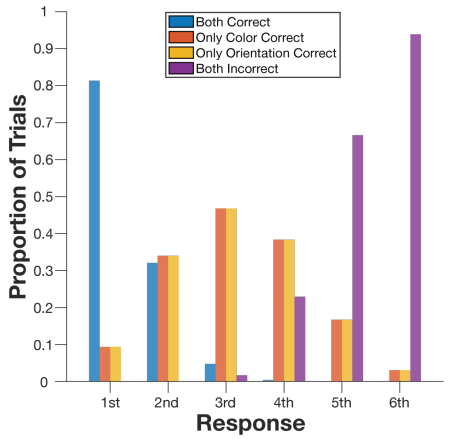
Pointer Model

Item-based storage with feature loss

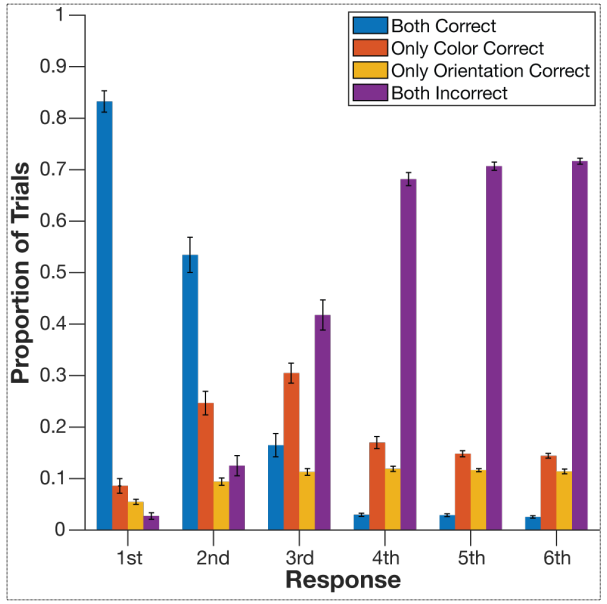


Independent Feature Model

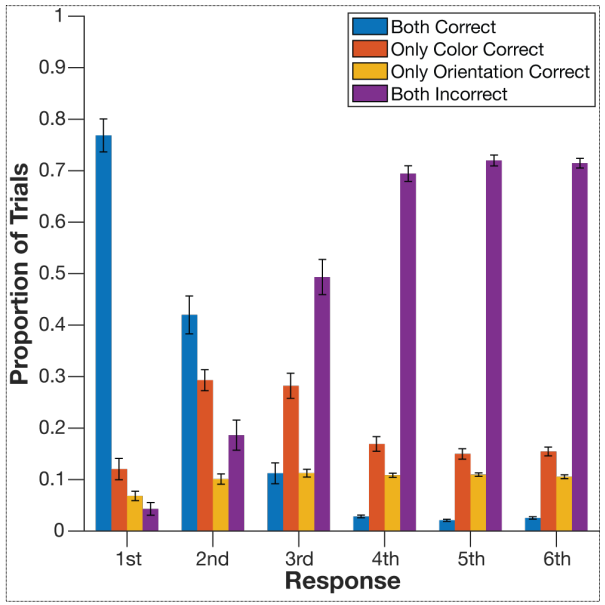
Feature storage independent of objecthood



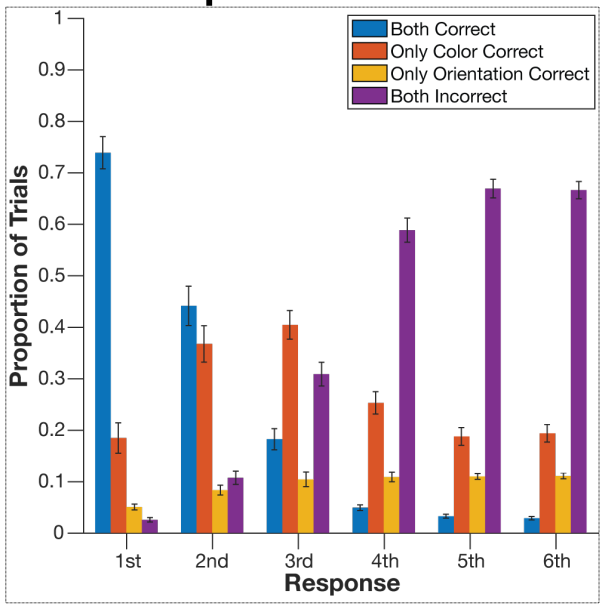
Experiment 1



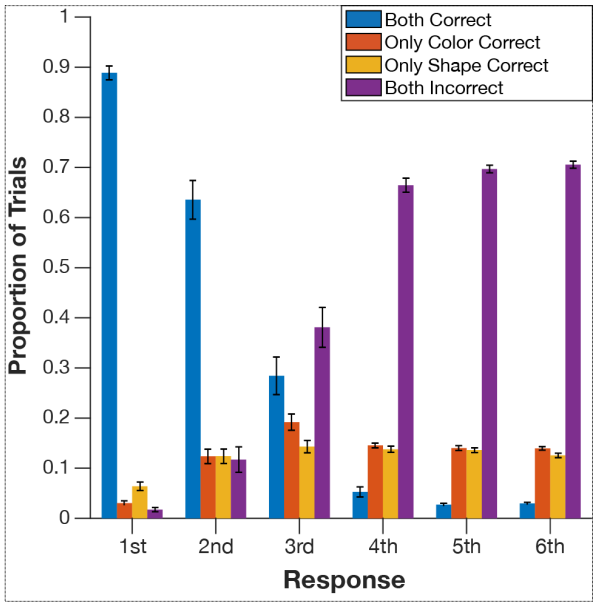
Experiment 2



Experiment 3

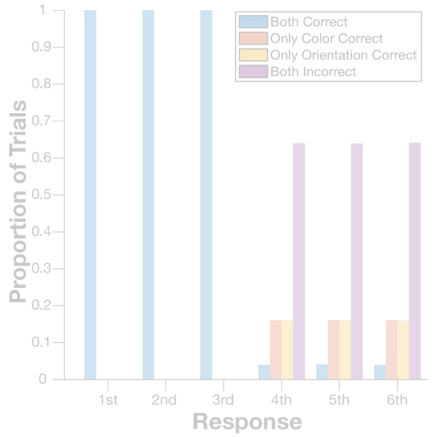


Experiment 4



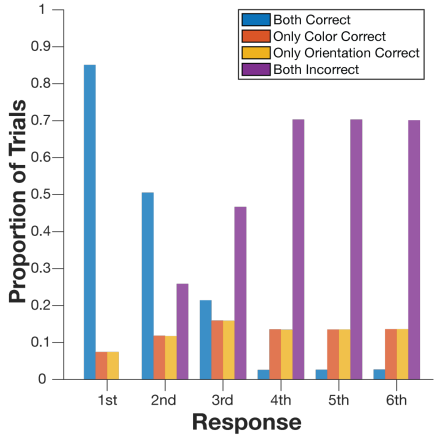
Strong Object Model

Accurate storage of three objects



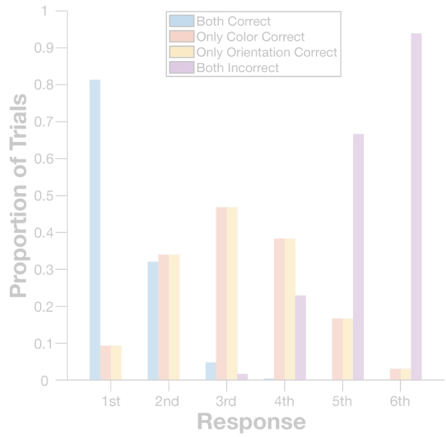
Pointer Model

Item-based storage with feature loss



Independent Feature Model

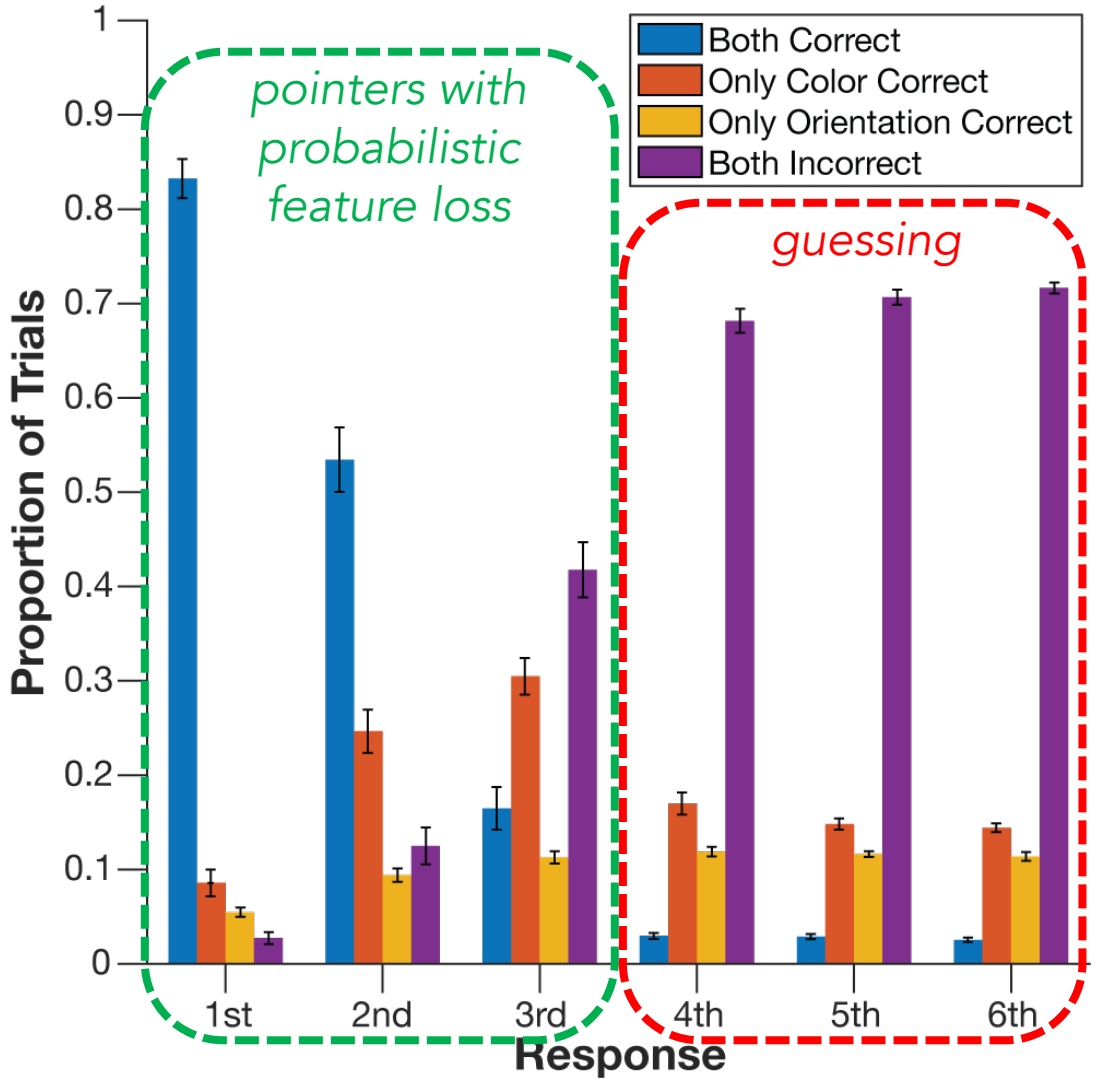
Feature storage independent of objecthood



A pointer model

- Pointers are a mechanism to maintain representations of objects through changes in its features
  - *FINSTs* (Pylyshyn, 1989)
  - *Object Files* (Kahneman et al., 1992)

- Not simply objects *or* features
  - We see object-based and feature-based phenomena in concert



Pylyshyn, Z. (1989). https://doi.org/10.1016/0010-0277(89)90014-0
Kahneman, D., Treisman, A., & Gibbs, B. J. (1992). https://doi.org/10.1016/0010-0285(92)90007-O
Thyer, W. et al. (2022). https://doi.org/10.1177/09567976221090923

What have we learnt from these projects?

- **Guess bands:**
  - We find clear evidence for guessing, in line with a discrete item limit model
  - But a continuous resources (*variable-precision*) model can still account for the pattern of data
    - With an *ad hoc* inclusion of priors
- **Conjunction whole-report:**
  - We see both object-based and feature-based phenomena occurring in concert
  - Working memory is not simply explained as objects or features, likely to be both!

How do we address the theory crisis?

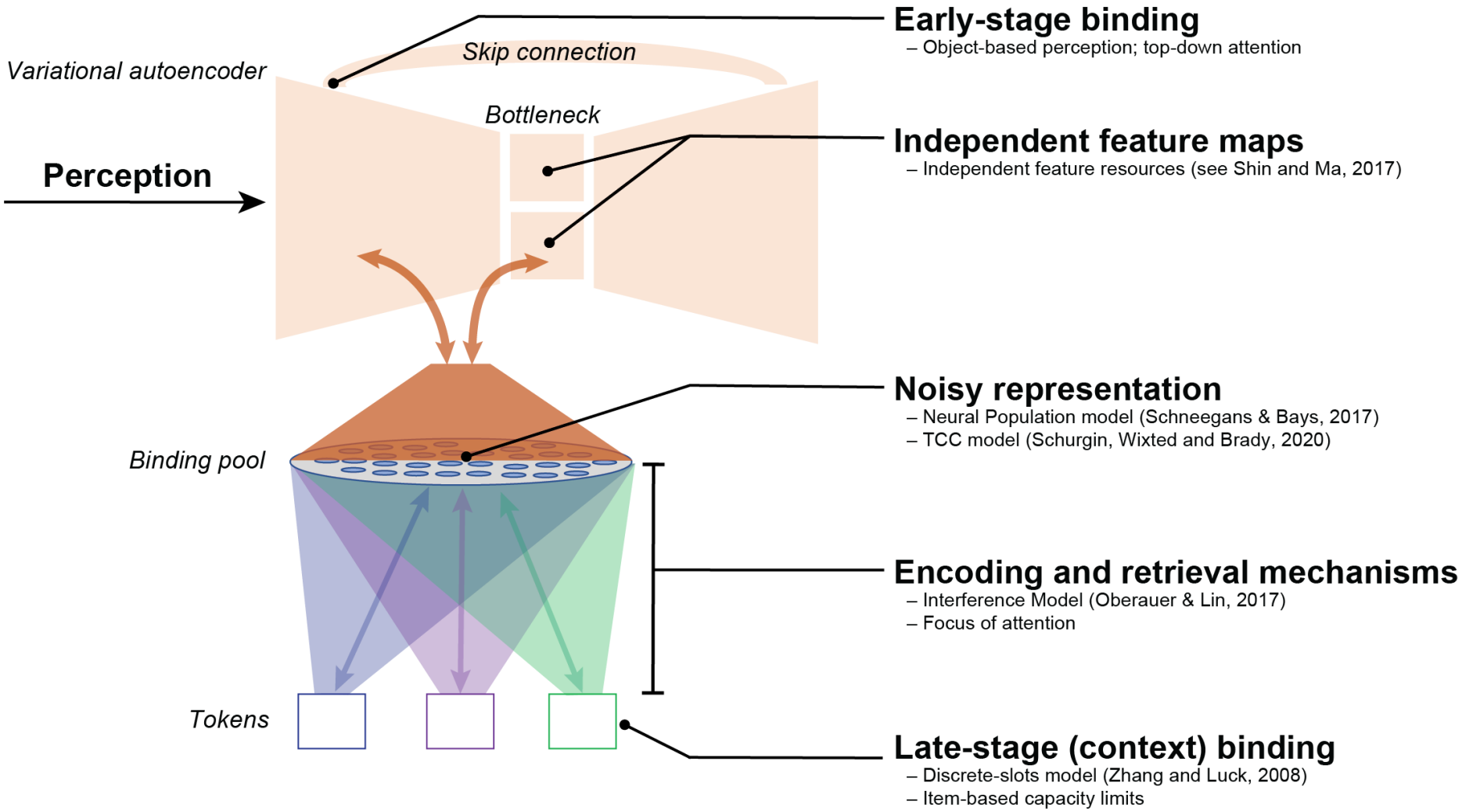
Can we bring these models into accordance?

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

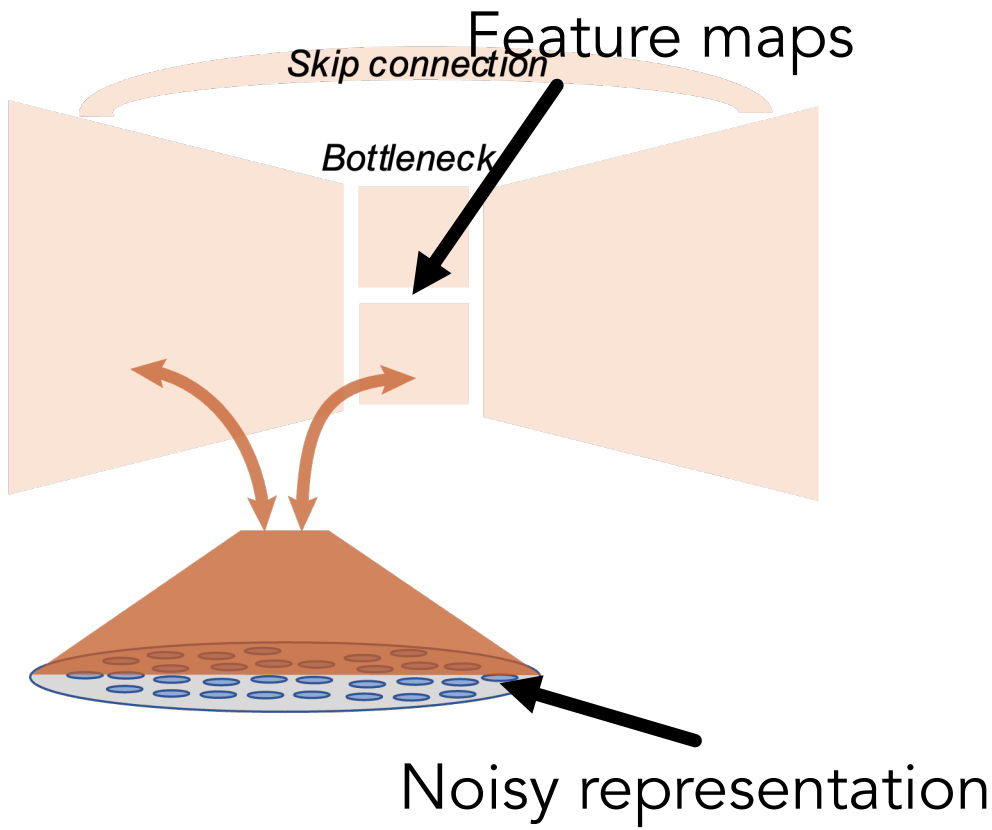


Hedayati, S., O'Donnell, R. E., & Wyble, B. (2022). A model of working memory for latent representations. *Nature Human Behaviour*, *6*(5), 709-719.

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)

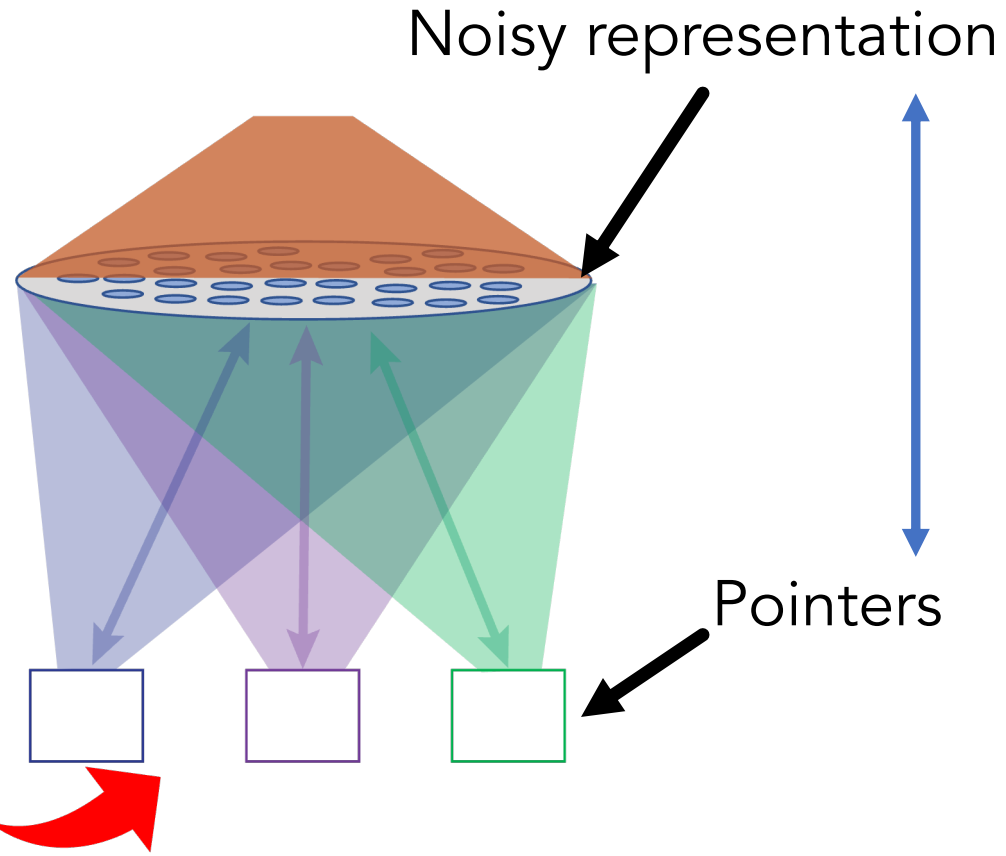


Bays, P. M. (2014). Noise in neural populations accounts for errors in working memory. *Journal of Neuroscience*, *34*(10), 3632-3645.
Schurgin, M. W., Wixted, J. T., & Brady, T. F. (2020). Psychophysical scaling reveals a unified theory of visual memory strength. *Nature human behaviour*, *4*(11), 1156-1172.
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)

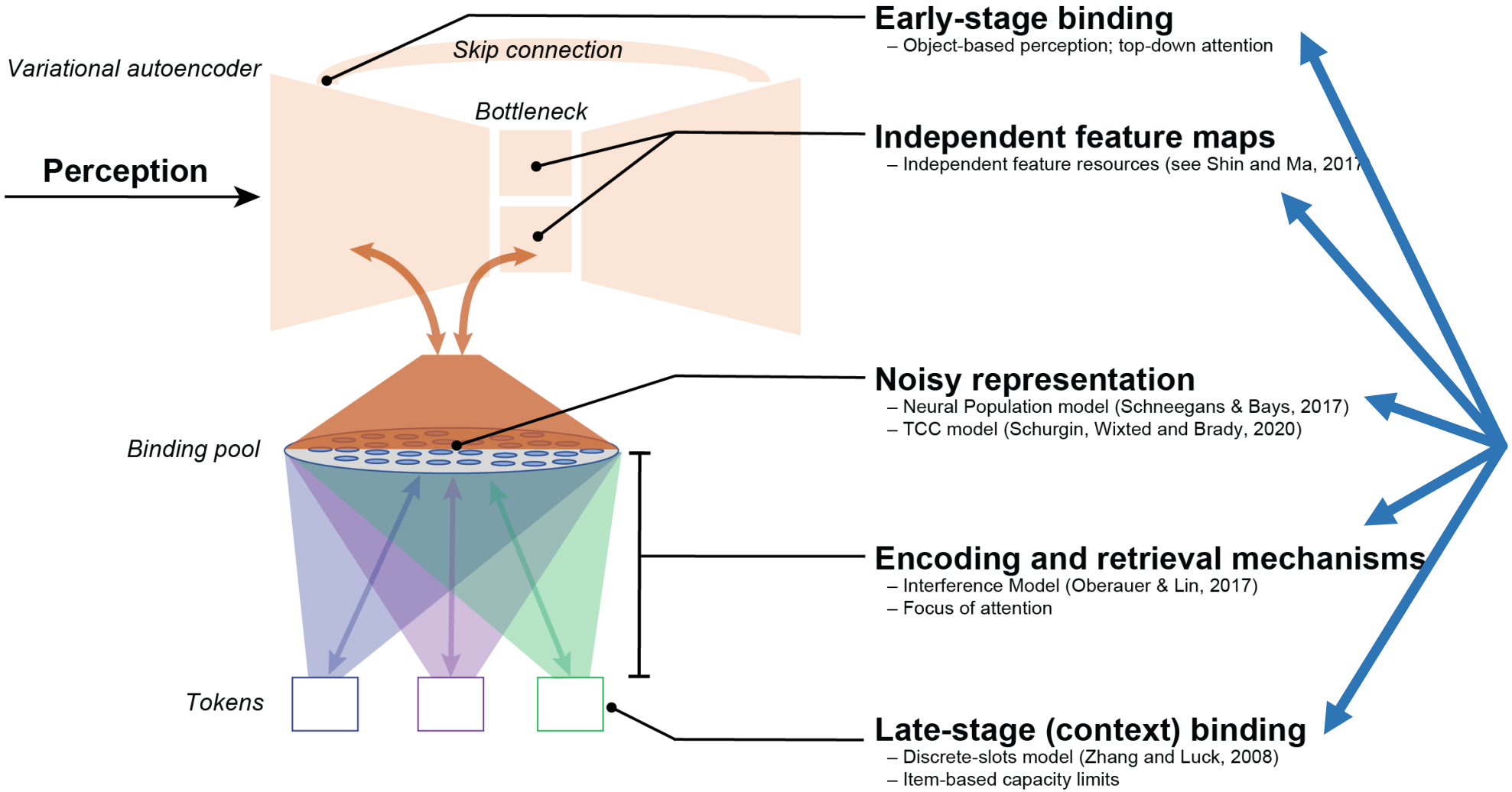
- Evidence for a neural signature that indexes VWM load and generalizes across feature content (Thyer et al., 2022; Balaban et al., 2019)



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

Exponential growth of scientific publications

Feeding a capitalistic academic system devalues the work.

Slow down, think again, test carefully.

Figure taken from arxiv.org on the number of submissions over time. https://arxiv.org/stats/monthly_submissions

Main messages

- Think carefully about how the theories connect to your tests
  - Are you sure the theories make a specific prediction?
  - Will your studies severely test the theory?

- Remember that memory is multi-faceted
  - Do you have a precise enough measure?
  - What inference can you make or what can you model in the system with your data?

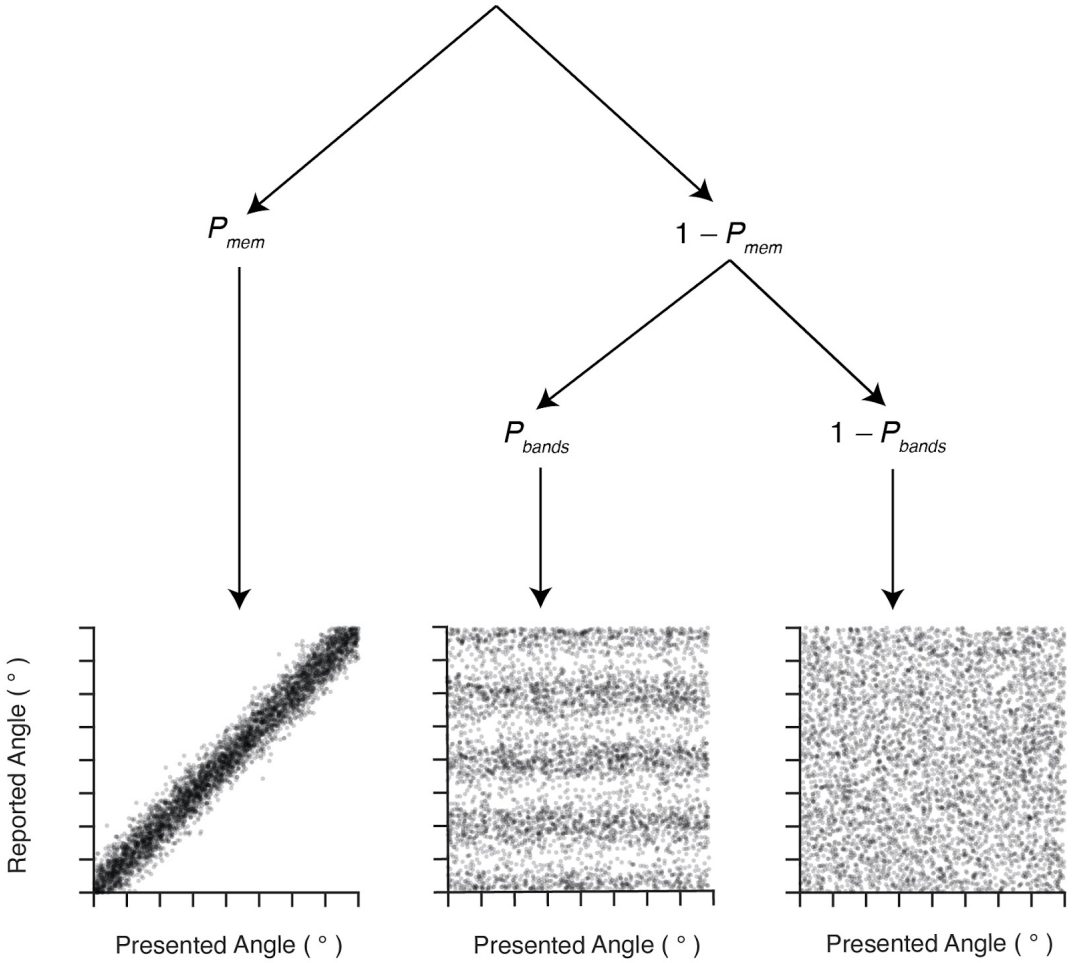
- My theory map can help you think about visual working memory:
  Ngiam, W. X. Q. (2023). Mapping visual working memory models to a theoretical framework. Psychonomic Bulletin & Review, 1-18.

Dr William Xiang Quan Ngiam *williamngiam.github.io* *wngiam@uchicago.edu*

Hidden track 1:
Are you sure those were *guesses*?

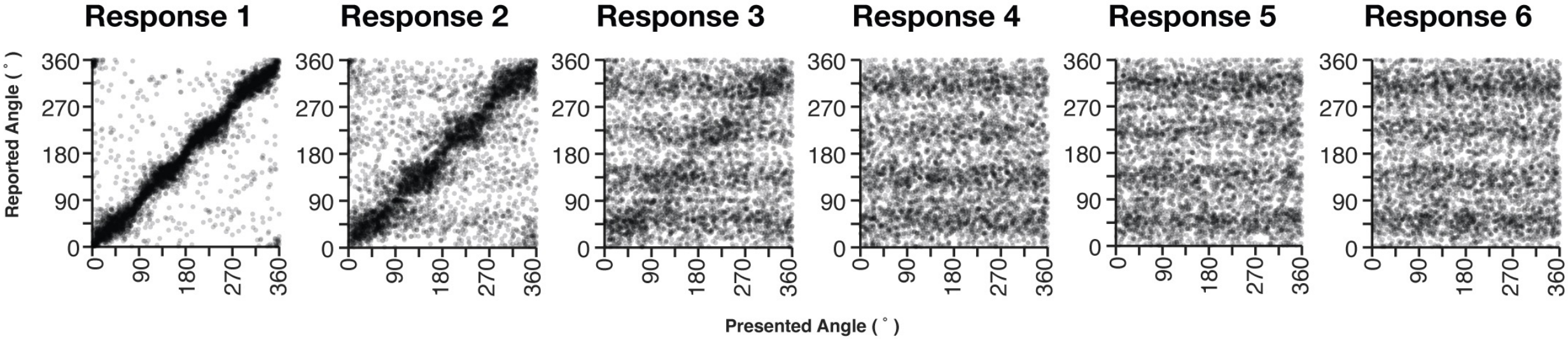
Formal model comparison

- Maximum likelihood estimation of the parameters for models with each possible permutation of the components:
  - Von Mises (a memory response)
    - Width of the Von Mises was a free parameter
  - Bands (a guess response)
    - Width of the bands was a free parameter
  - Uniform (a random response)
- 100 replicates with a maximum of 10000 iterations
  - Compared on the Bayesian Information Criterion (BIC)



Experiment 1 model comparison

- At the **aggregate level:**
  - For the first three responses, Von Mises + Guess Bands was the best-fitting model (ΔBIC < 9).
  - For the last three responses, Von Mises + Guess Bands + Uniform was the best-fitting model (ΔBIC > 57)



Estimated prevalence of responses

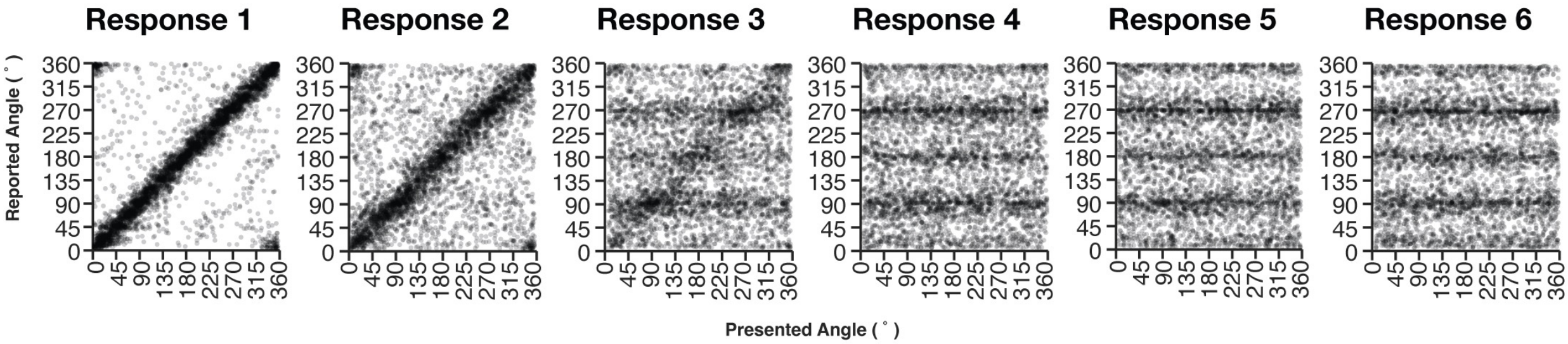
- Parameter estimates from Von Mises + Guess Bands + Uniform model

Response	Memory	Guess Bands	Uniform
1st	90.59% ± 0.57%	9.41% ± 1.15%	0% ± 0.58%
2nd	66.03% ± 1.68%	33.97% ± 2.20%	0% ± 0.52%
3rd	20.37% ± 0.63%	46.64% ± 12.16%	32.99% ± 11.53%
4th	0.19% ± 0.09%	41.96% ± 8.29%	57.85% ± 8.20%
5th	0.30% ± 0.12%	35.78% ± 4.53%	63.92% ± 4.41%
6th	0.39% ± 0.12%	39.12% ± 6.25%	60.49% ± 6.13%

- Memory responses are constrained to the first three responses
- Substantial prevalence of 'guess band' responses in later responses

Experiment 2 model comparison

- At the **aggregate level**:
  - For the first response, Von Mises + Uniform was the best-fitting model (ΔBIC = 8).
  - For the last four responses, Von Mises + Guess Bands + Uniform was the best-fitting model (ΔBIC > 24 from 3rd response onward)



Estimated prevalence of responses

- Parameter estimates from Von Mises + Guess Bands + Uniform model

Response	Memory	Guess Bands	Uniform
1st	87.84% ± 0.00%	0.64% ± 0.00%	11.52% ± 0.00%
2nd	64.13% ± 1.18%	2.08% ± 0.90%	33.79% ± 2.08%
3rd	21.07% ± 0.61%	37.26% ± 6.25%	41.67% ± 5.65%
4th	0.31% ± 0.11%	48.10% ± 6.02%	51.59% ± 5.91%
5th	0.21% ± 0.11%	48.70% ± 4.70%	51.09% ± 4.58%
6th	0.25% ± 0.11%	47.22% ± 4.35%	52.53% ± 4.24%

- Memory responses are constrained to the first three responses
- Substantial prevalence of 'guess band' responses in later responses

Formal model comparison on individual data

- Experiment 1
  - In early responses, the Von Mises + Uniform (M1) model best fits most participants' data
  - In later responses, the Guess Bands only (M4) model best fits most participants' data

	M1	M2	M3	M4	M5	M6
1st	**28**	-	-	-	10	2
2nd	**19**	-	1	2	18	-
3rd	**14**	-	1	2	13	-
4th	6	-	-	**30**	4	-
5th	5	2	2	**25**	6	-
6th	6	1	2	**23**	8	-

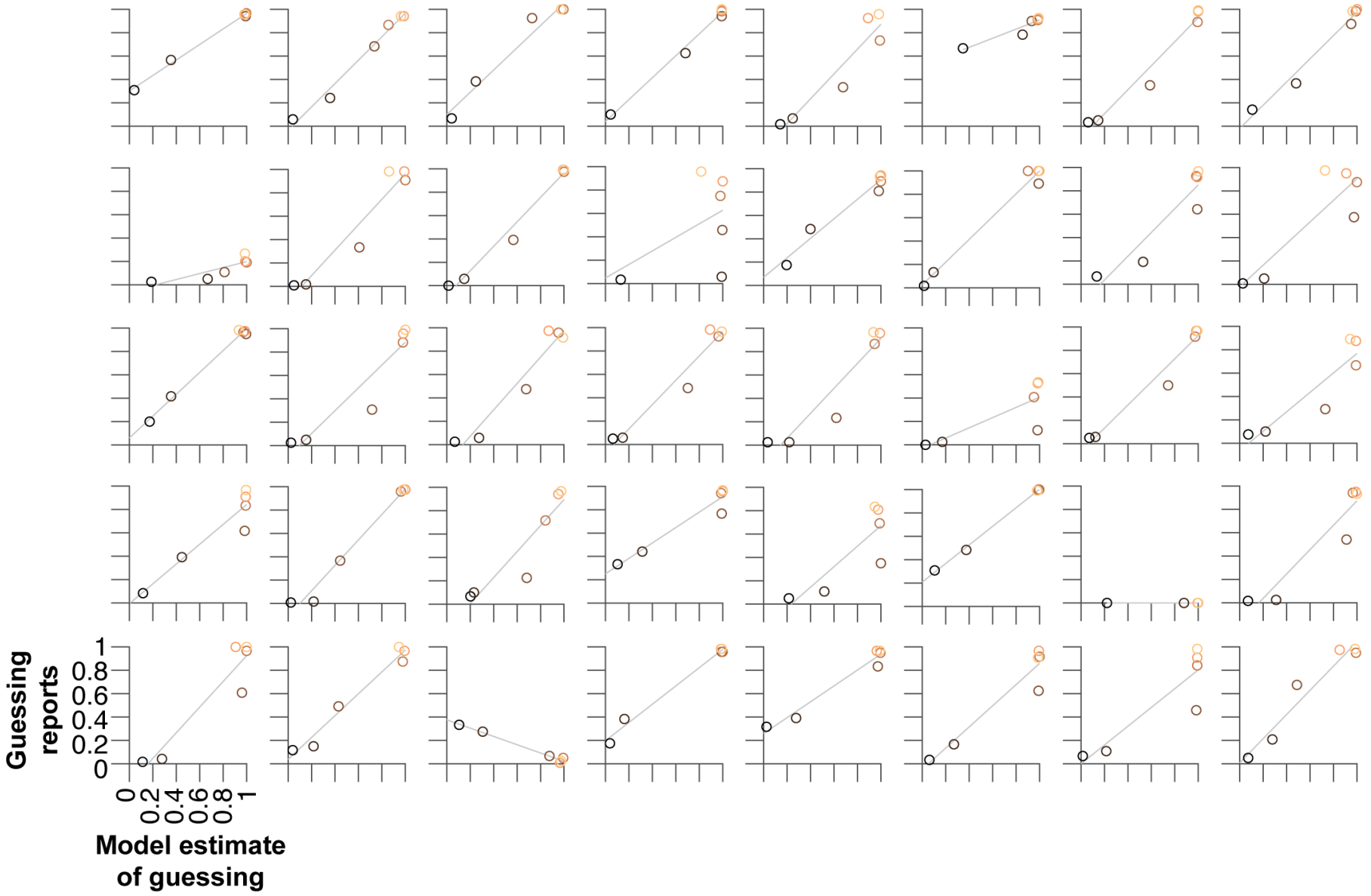
Formal model comparison on individual data

- Experiment 2
  - In early responses, the Von Mises + Uniform (M1) model best fits most participants' data
  - In later responses, the Guess Bands only (M4) model best fits most participants' data

	M1	M2	M3	M4	M5	M6
1st	**23**	-	1	-	4	2
2nd	**17**	-	3	-	10	-
3rd	4	4	5	7	**10**	-
4th	4	7	5	**9**	5	-
5th	5	**11**	1	**11**	2	-
6th	1	5	3	**16**	5	-

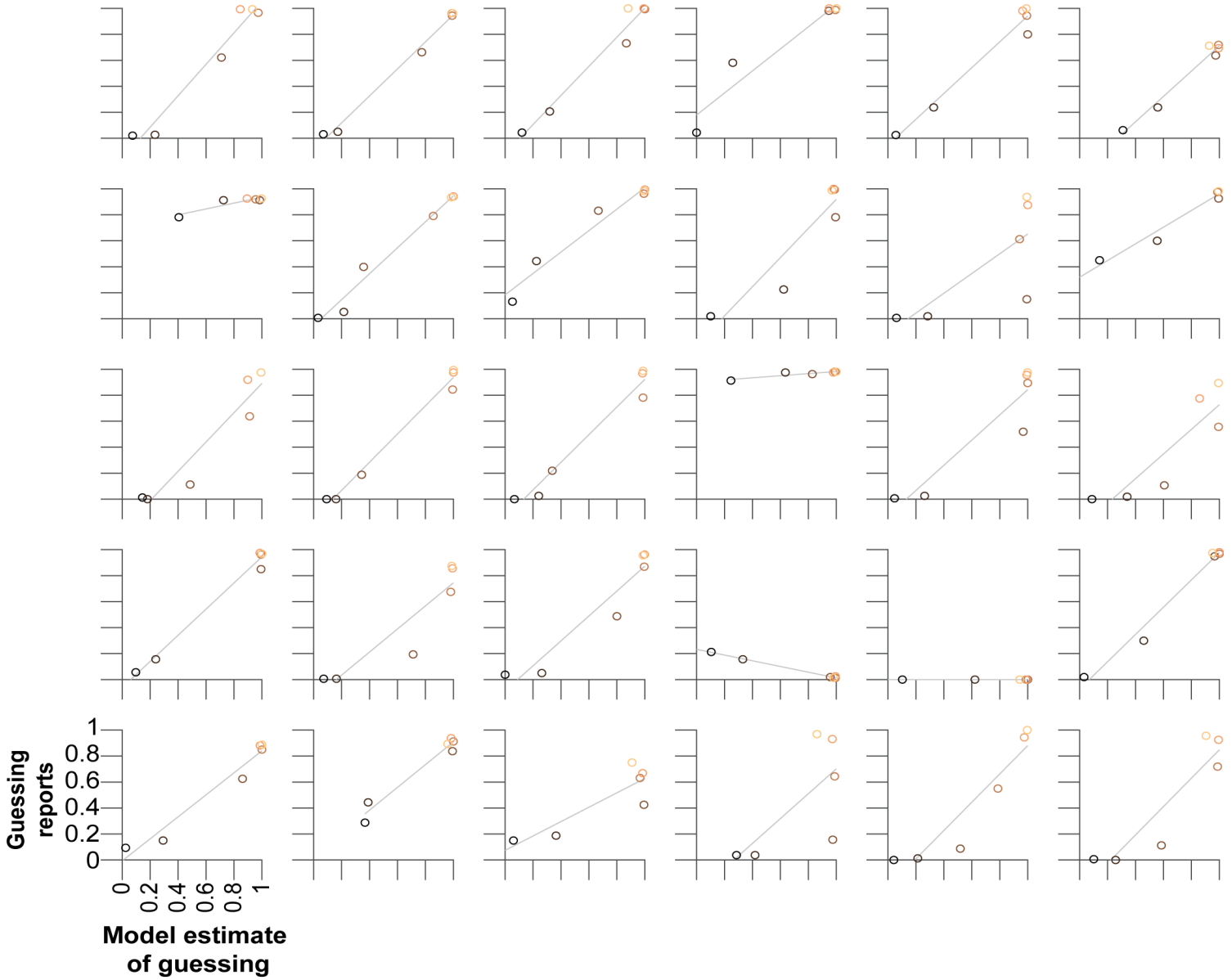
Self-reports of guesses match model estimates

- Experiment 1 (background condition)



Self-reports of guesses match model estimates

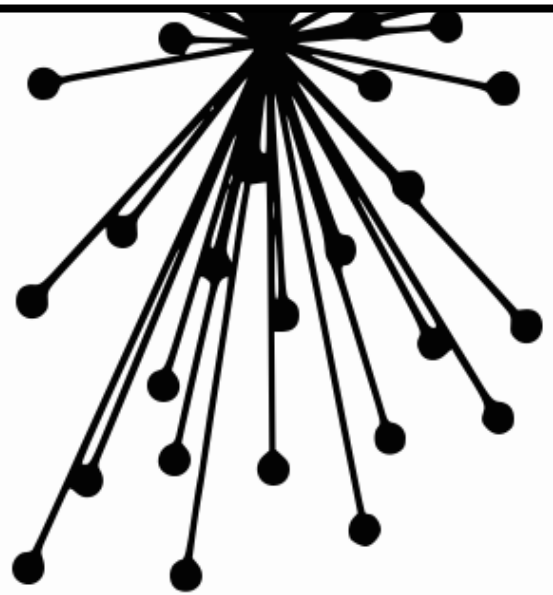
- Experiment 2



Hidden track 2:
How do we fix *science*?

This hasn't worked.

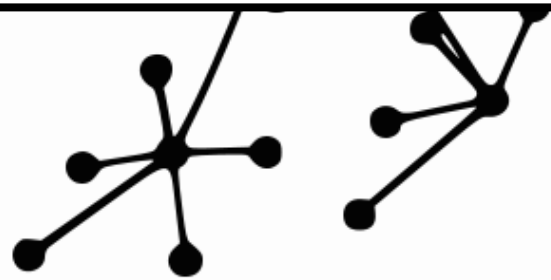
e.g. I've seen little support from my department in the past 4 years.



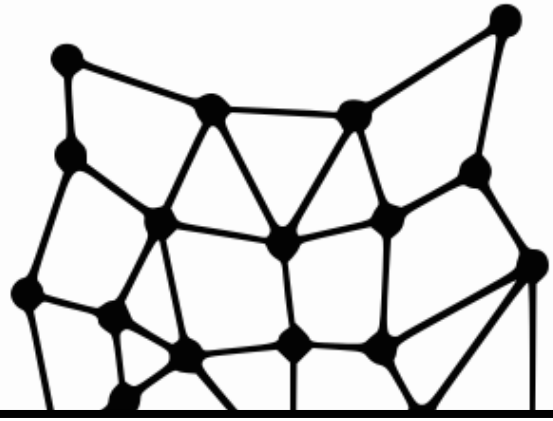
Centralized

This could work...

But reform is siloed, and would not spread across the network.



Decentralized

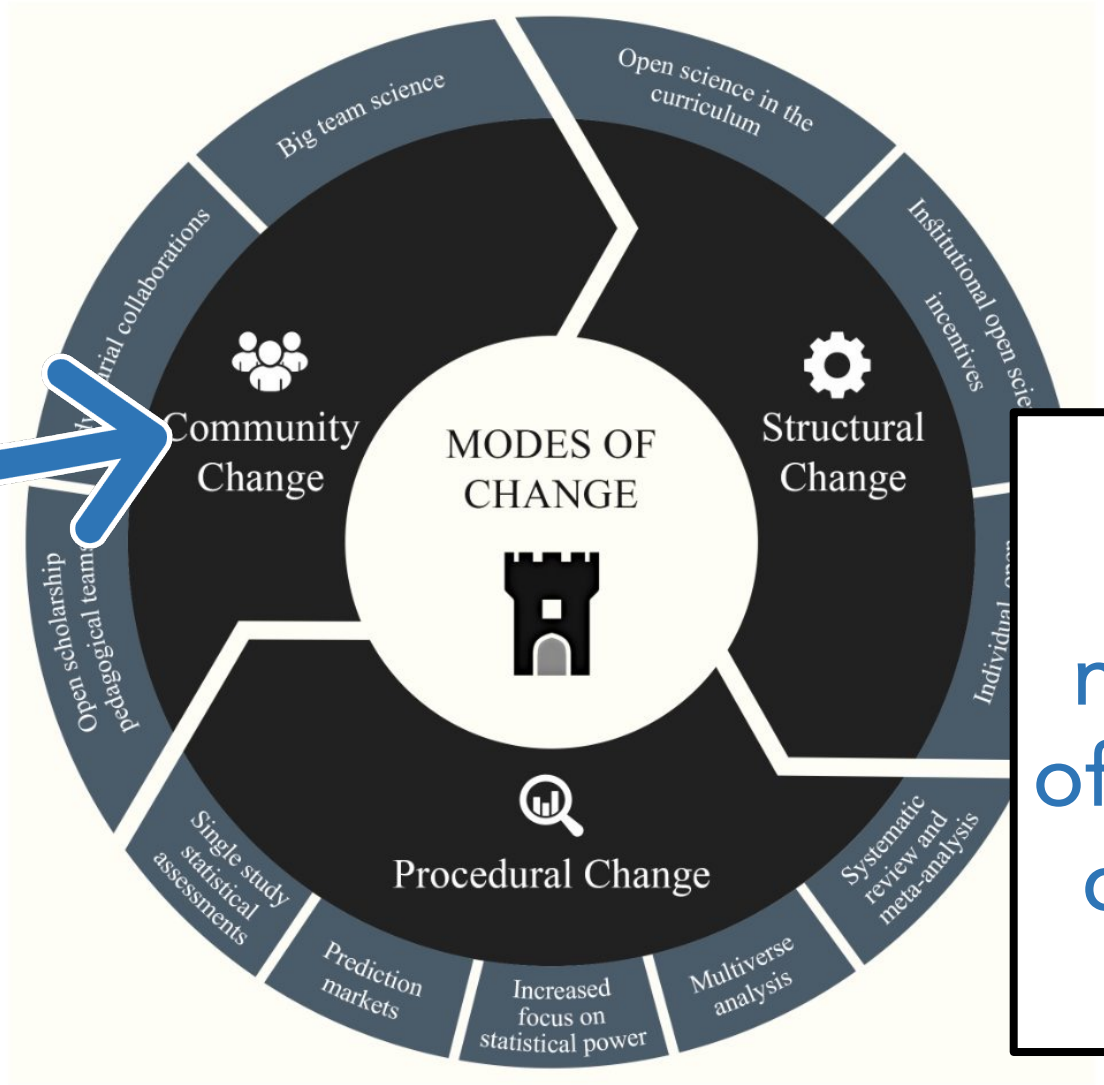


This is best.

We need reform at many levels, and changes to spread through the network.

Distributed

Not doing anything adds resistance to changes and reforms. It calcifies existing structures.



ReproducibiliTea

It is my firm belief that the next generation of researchers will change science for the better

Fig. 1 Modes of change towards scientific credibility. This figure presents an overview of the three modes of change proposed in this article: structural change is often evoked at the institutional level and expressed by new norms and rules; procedural change refers to behaviours and sets of commonly used practices in the research process; community change encompasses how work and collaboration within the scientific community evolves.

Korbmacher, M., Azevedo, F., Pennington, C., Hartmann, H., Pownall, M., Schmidt, K., ... & Evans, T. (2023). The replication crisis has led to positive structural, procedural, and community changes. *Communications Psychology*.

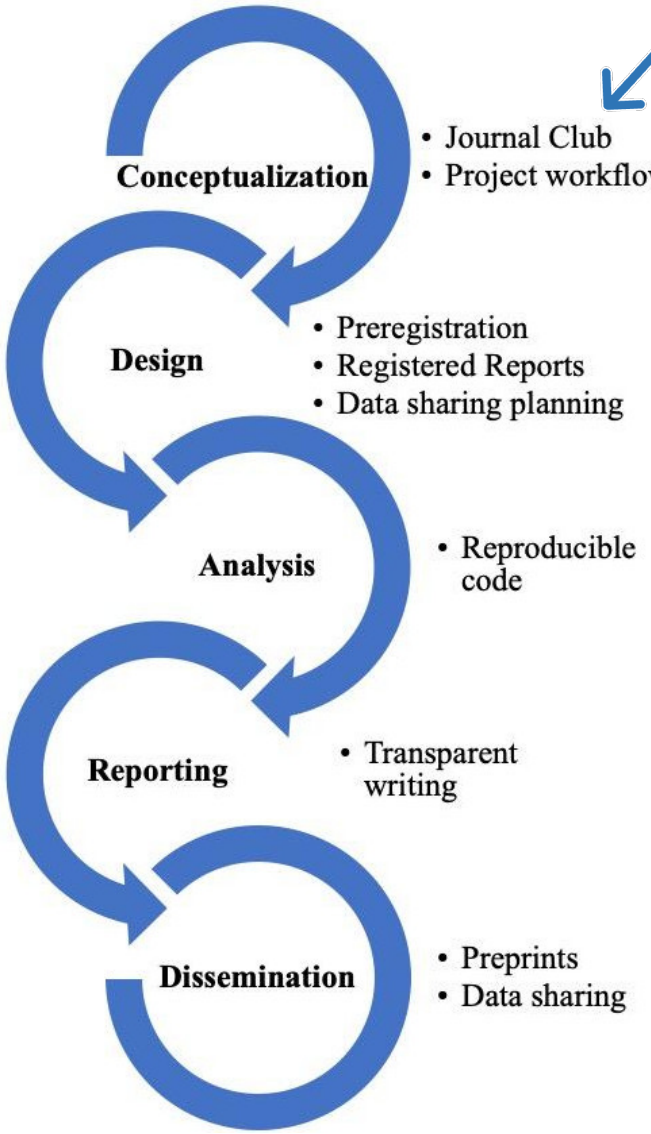
You don’t have to do this on your own

- Only one hour or so out of your week
- Form community with your fellow junior scientists from otherwise siloed areas in the department
- Develop and get advice on your research and science
- Have the hidden curriculum of academia and science revealed

Where do I start?

- Open Science is not "all or nothing"
  - These are research skills that take time to develop!

- Some easy Open Science practices to adopt:
  - Post free copies of published articles / deposit preprints of all manuscripts
  - Publish in open access venues
  - Publicly share data and materials
  - Preregister studies



Kathawalla, U. K., Silverstein, P., & Syed, M. (2021). Easing into open science: A guide for graduate students and their advisors. *Collabra: Psychology*, *7*(1).
McKiernan, E. C., Bourne, P. E., Brown, C. T., Buck, S., Kenall, A., Lin, J., ... & Yarkoni, T. (2016). Point of view: How open science helps researchers succeed. eLife, 5, e16800.

ReproducibiliTea Introductory Reading List

These are our recommendations for the papers to cover in the first term of your new ReproducibiliTea journal club! These ten papers were selected to provide an overview of the reproducibility crisis and introduction to the many aspects of Open Science. They are separated by themes that your journal club may choose to explore in further detail in following meetings! We have also provided a summary, keywords and online resources to help inform your discussions.

Order	Block	Paper	Summary	Keywords	Resources
1	The 'issues' that lead to the reproducibility crisis	Ioannidis JPA (2005). Why most published research findings are false. PLoS Med 2(8): e124. https://doi.org/10.1371/journal.pmed.0020124	**Defining the issue.** By simulating at various levels of statistical power, across different pre-study odds, the accumulation of significant results is shown to be potentially false positives predominantly. The paper introduces concepts like the positive predictive value and how it is related to the *p*-value, and how important having high statistical power is for the rigor of research.	*p*-values, positive predictive values, false positives, statistical power	Summary video (by William Ngiam): https://www.youtube.com/watch?v=C7N_-XanpTI
2		Smaldino, P. E., & McElreath, R. (2016). The natural selection of bad science. Royal Society open science, 3(9), 160384. https://doi.org/10.1098/rsos.160384	**The myth of self-correction.** Estimates of statistical power historically in science appears to be extremely low. In addition to that, due to publication bias (the view that positive results are more likely to be published) and the incentives to publish, simulations suggest that a high false-discovery rate is 'naturally selected' for and that replications are ineffective at correcting that rate.	statistical power, replication	Summary video (by William Ngiam): https://www.youtube.com/watch?v=EdLDE2Y4exM
3		Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant. Psychological Science, 22(11), 1359–1366. https://doi.org/10.1177/0956797611417632	**The problem of analytic flexibility.** A demonstration of how decisions made by researchers in statistical analysis, such as dropping conditions or adding observations after a non-significant test, can easily produce a false positive result.	analytic flexibility, researcher degrees of freedom, questionable research practices	Summary video (by William Ngiam): https://www.youtube.com/watch?v=bf3GqyBRgzY
4	The extent of the 'issues'	John, L. K., Loewenstein, G., & Prelec, D.	**The prevalence of questionable research practices.** With an	questionable research practices	
5			reform.		/v=IDTMWKDghLM
6	Perspectives on the reproducibility crisis	Vazire, S. (2018). Implications of the credibility revolution for productivity, creativity, and progress. Perspectives on Psychological Science, 13(4), 411-417. https://doi-org/10.1177%2F1745691617751884	**The credibility revolution.** A reframing of the 'reproducibility crisis' that highlights the scientific reforms that have occurred with the Open Science movement, and their potential impacts on the productivity, creativity and progress of scientists.	credibility revolution, commentary, summary	Presentation by Simine Vazire at OSC 2019: https://www.youtube.com/watch?v=Yf1Ovx-OixE
7		Yarkoni, T. (2018), Not its not The Incentives - it's you. Yarkoni Blog - [citation needed]: https://www.talyarkoni.org/blog/2018/10/02/no-its-not-the-incentives-its-you/	**Dealing with the Incentives.** A blogpost arguing that the responsibility for reproducible science rests with the individual, and that the Incentives are not a good reason to be absolved of that responsibility	incentives, commentary	
8	Getting started with Open Science	Kathawalla, U. K., Silverstein, P., & Syed, M. (2021). Easing into open science: A guide for graduate students and their advisors. Collabra: Psychology, 7(1). https://doi.org/10.1525/collabra.18684	**Easing into Open Science.** A very accessible guide for graduate students (and their advisors) on some of the different ways to engage with the reproducibility movement. They are given difficulty ratings (easy, medium or difficult) and potential worries are also addressed.	early-career researchers, guide, introductory, pre-registration	Presentation by Priya Silverstein at RIOT Science Club: https://www.youtube.com/watch?v=owJaD3UiseQ
9		Munafò, M. R., Nosek, B. A., Bishop, D. V. M., Button, K. S., Chambers, C. D., Percie Du Sert, N., Simonsohn, U., Wagenmakers, E. J., Ware, J. J., & Ioannidis, J. P. A. (2017). A manifesto for reproducible science. Nature Human Behaviour, 1(1), 1–9. https://doi.org/10.1038/s41562-016-0021	**A manifesto for reproducible science.** A general overview of the goals of various reproducibility measures and how they can be implemented.	guide, reproducibility	
10		Crüwell, S., van Doorn, J., Etz, A., Makel, M. C., Moshontz, H., Niebaum, J. C., ... & Schulte-Mecklenbeck, M. (2019). Seven easy steps to open science. Zeitschrift für Psychologie. http://dx.doi.org/10.1027/2151-2604/a000387	**Where to next?** An annotated reading list of papers from seven topics: open access, open data, preregistration, reproducible analyses, replications and teaching open science in an attempt to make those practices more understandable and actionable for readers	transparency, meta-science	

ReproducibiliTea Reading List on Theory in Psychological Science

One precursor to the reproducibility crisis in psychology has been the haste to conduct empirical research, rather than rigorously develop theory and its connection to the research. These ten papers were selected to provide an introduction to theoretical psychology. They are separated by themes that your journal club may choose to explore in further detail in following meetings! We have also provided a brief summary, keywords and additional online resources to help inform your discussions.

Order	Block	Paper	Summary	Keywords	Resources
1	What is a theory?	Fried, E. I. (2020). Theories and models: What they are, what they are for, and what they are about. Psychological Inquiry, 31(4), 336-344. https://doi.org/10.1080/1047840X.2020.1854011	**What is a theory?** An overview on the role of theories and models in science, including a brief commentary on the weakness of theories in the psychological sciences and how to make them better.	introductory, theory development	Eiko Fried on "Theory building and testing in psychological research" for the RIOT Science Club: https://youtu.be/vB1Hk3c-lZY
2	Does psychology have a theory problem?	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 1978, Vol. 46, 806-834. https://www3.nd.edu/~ghaeffel/Meehl(1978).pdf	**The lack of theory development in psychology.** An astute criticism of the excessive use of null hypothesis significance testing in 'soft psychology' that left psychological theories lacking "the cumulative character of scientific knowledge because they tend to be neither refuted nor corroborated, but instead merely fade away as people lose interest."	NHST, statistical testing, scientific inference	A video recording of the first lecture by Paul Meehl in his course on philosophical psychology from 1989, where he contrasts the role of theory in the 'hard sciences' like physics and the 'soft science' of psychology. https://youtu.be/AEPbzCTneDs
3		Klein, S. B. (2014). What can recent replication failures tell us about the theoretical commitments of psychology?. Theory & Psychology, 24(3), 326-338. https://doi.org/10.1177/0959354314529616	**A crisis in replication or beyond?** Determining success or failures of replications necessitates that theories be well-specified – clearly defining the relation between theory and prediction by linking rigorously established constructs to physical observations and detailing the essential conditions of experiments.	reproducibility crisis, replications, theory development	A personal commentary by Daniel Nettle on the pretense of having a theory in psychology: "Theories and models are not the only fruit" https://leonidtiokhin.medium.com/theories-and-models-are-not-the-only-fruit-a05c7cf188f6
4		Scheel, A. M., Tiokhin, L., Isager, P. M., & Lakens, D. (2021). Why hypothesis testers should spend less time testing hypotheses. Perspectives on Psychological Science, 16(4), 744-755. https://doi.org/10.1177/1745691620966795	**Are we ready to test?** Psychologists have been trained with a recipe: the hypothetico-deductive method – formulate a hypothesis from theory, devise a study to test the hypothesis, collect and analyze data, and finally evaluate whether there is support for or against the theory. However, without the groundwork strengthening the 'derivation chain that links theory to hypothesis test, the confirmatory testing that is prized by the recent reform movement may be premature.	exploratory versus confirmatory, derivation chain	Anne Scheel on "Equivalence testing for psychological research" for the RIOT Science Club https://youtu.be/T9pzORPTXFU
5					ynote presentation by Olivia t on "What makes a good theory, ow do we make a theory good?" ://youtu.be/67X0TpnQeO0
6	Improve psychological theory	Psychological Science, 16(4), 789-802. https://doi.org/10.1177/1745691620970585	theories as abstract constructs are formalized, and underlying intuitions and predictions are made open and transparent.		eo recording of a talk by Olivia t and Andrea Martin on their r "How computational modeling can force theory building in psychological science". https://youtu.be/8Aa9_6ahO48
7		Maatman, F. O. (2021). Psychology's Theory Crisis, and Why Formal Modelling Cannot Solve It. https://psyarxiv.com/puqvs/	**Formal theories are helpful but first be determined.** The cause of the theory crisis stems from tests of experiments not being specific enough as to support only one theory and falsify all other alternatives, and many psychological theories containing auxiliary assumptions such that the theories are not severely tested. Better methods that force precise and unlikely predictions from theories will solve the core issue, not necessarily formal modeling alone.	proto-theory, formal theory, theory building, theory specification	A Twitter thread by Freek Maatman (@psychedfreek) summarizing their paper. https://twitter.com/psychedfreek/status/1414982603082506242
8	Taking steps to improve psychological theory	Flake, J. K., & Fried, E. I. (2020). Measurement schmeasurement: Questionable measurement practices and how to avoid them. Advances in Methods and Practices in Psychological Science, 3(4), 456-465. https://doi.org/10.1177/2515245920952393	**Better measures to inform theory building.** Developing and testing theories requires construct measures to be scrutinized and valid. Echoing questionable research practices, questionable measurement practices (e.g. the arbitrary summing of subscales) are defined and a list of questions are provided to help the researcher promote the validity of their measures.	guide, measurement, transparency, construct validity	Jessica Flake on "Measurement schmeasurement: Questionably measurement practices and how to avoid them" for the RIOT Science Club: https://youtu.be/Cq6n7AS_r8w
9		van Rooij, I., & Blokpoel, M. (2020). Formalizing verbal theories: A tutorial by dialogue. Social Psychology, 51(5), 285. https://doi.org/10.1027/1864-9335/a000428	**Formalizing verbal theories.** A guide to translating verbal theories into formal theories starting with basic mathematical definitions and notation before a toy example of building formal theories presented through multiple dialogues between fictional Dr Verbal and Dr Formal.	guide, theory building, formal modeling	Smaldino, P. E. (2020). How to translate a verbal theory into a formal model. Social Psychology, 51(4), 207.https://osf.io/preprints/metaarxiv/n7qsh
10	Are we ready to test our theories?	Oberauer, K., & Lewandowsky, S. (2019). Addressing the theory crisis in psychology. Psychonomic bulletin & review, 26(5), 1596-1618. https://doi.org/10.3758/s13423-019-01645-2	**Discovery-oriented research versus theory-testing research.** A critical and comprehensive revisit of the reproducibility crisis and proposed solutions, such as preregistration, formal modeling and stricter statistical standards. A distinction is made between two paths: discovery-oriented research, where it is accepted that theory cannot yet lead to strong inferences and necessitating empirical standards through direct replication, versus theory-testing research, where theories are formulated as precisely as possible as to close the gap between theory and hypothesis.	reproducibility crisis, preregistration, statistical inference, formal modelling, exploratory versus confirmatory, theory specification	Eiko Fried contextualizes and summarizes the field of theoretical psychology in his blogpost "On Theory". The inspiration for this resource, and a good place to inform where to go next to continue learning about psychological theory. https://eiko-fried.com/on-theory/

Created by William Xiang Quan Ngiam

On my website, under "reading lists"

ORIGINAL ARTICLE

PSYCHOPHYSIOLOGY SPR WILEY

Estimating the statistical power to detect set-size effects in contralateral delay activity

William X. Q. Ngiam[1] | Kirsten C. S. Adam[2] | Colin Quirk[1] | Edward K. Vogel[1] | Edward Awh[1]

PsyArXiv Preprints My Preprints Submit a Preprint Search Donate

Mapping visual working memory models to a theoretical framework

AUTHORS
William X. Q. Ngiam

It's not either/or – your goals can include improving science while conducting empirical research.

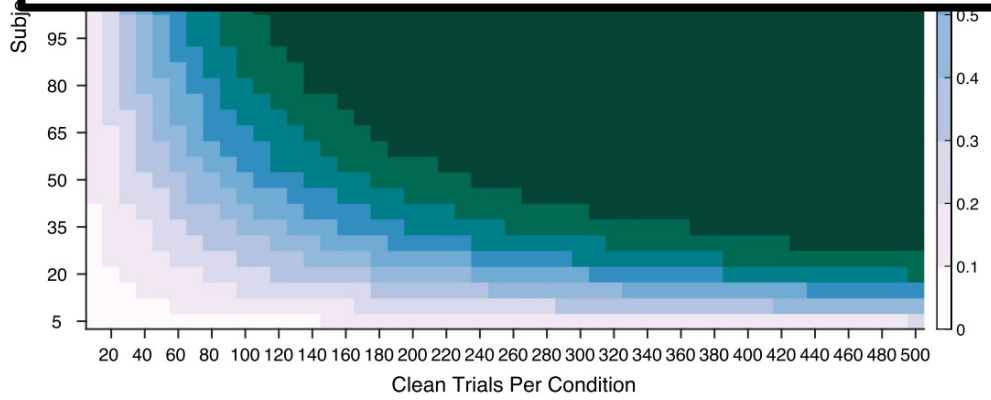


FIGURE 6 Simulated statistical power for observing a significant difference in CDA amplitude between set sizes 2 and 4 beyond the bounds of the Hakim et al. (2019) dataset

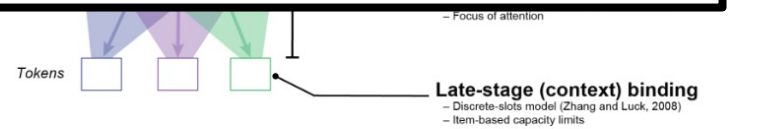


Figure 1. A simplified schematic of the Memory for Latent Representations (MLR) model architecture (Hedayati et al., 2022) with visual working memory phenomena and current models mapped on to its components: the variational autoencoder (VAE), the binding pool, and the tokens. This theory map aims to provide a coherent framework within which to organize visual working memory phenomena and discuss the relevant explanatory models. As such, the compatibility or inconsistencies between models can be better identified, and subsequently tested. For example, one could use a working definition for the noisy representation in VWM as the noise held in the pattern of neuron activity in the binding pool that follows a summation of information from various perceptual sources.

This seems like a lot more work…

- Rigor
- Saves
  - Ea
  - Cr
- Redu
  - Ca
  - Be
    rec
- Being
  resea

ater!

n

What is your goal in science?

Do you want to be beholden to the *Incentives*?

Or do you want to pave the way towards better understanding?

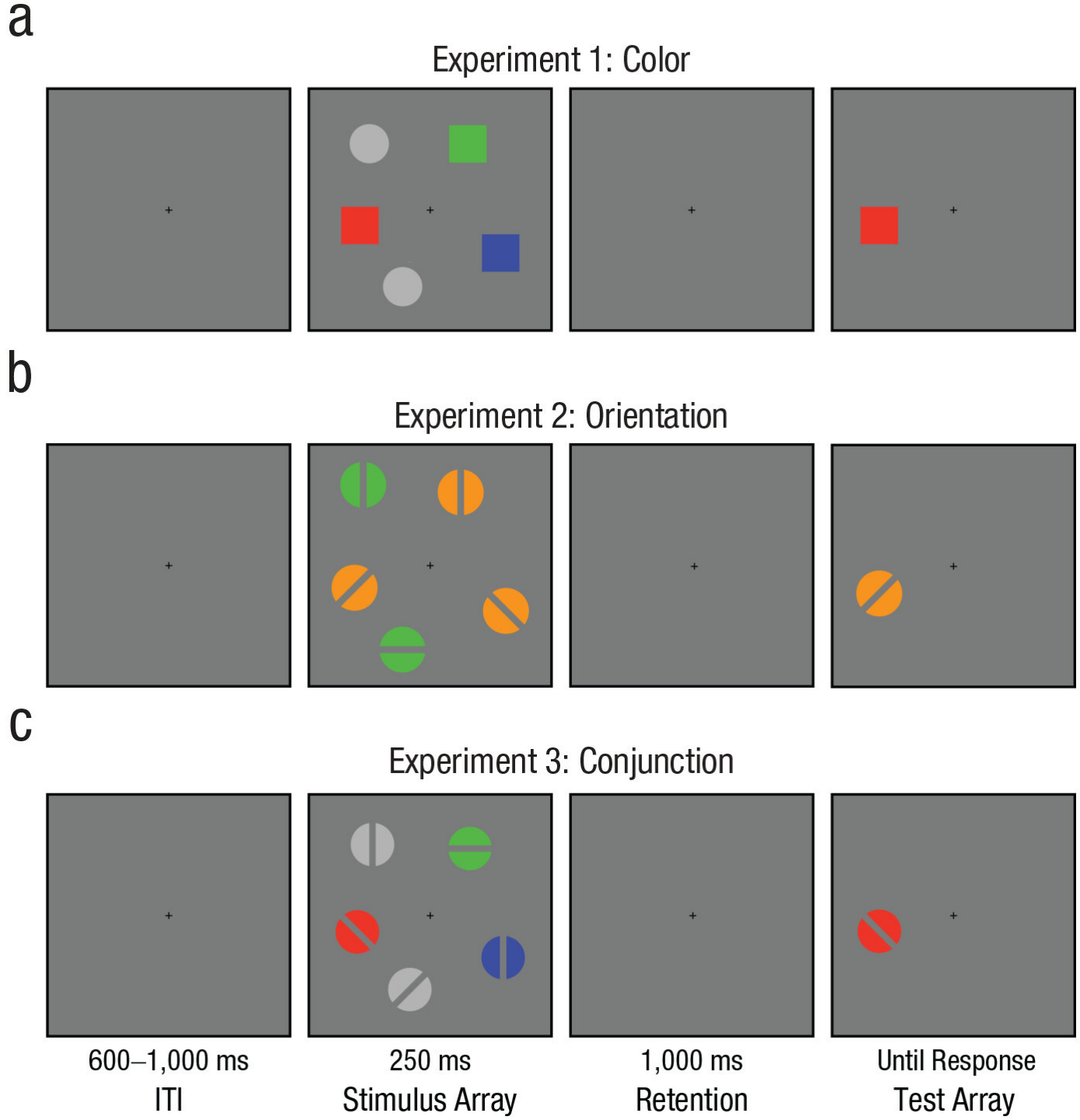
Yarkoni, T. (2018). It's not the Incentives, it's you. https://www.talyarkoni.org/blog/2018/10/02/no-its-not-the-incentives-its-you/
Hostler, T. (2023). The Invisible Workload of Open Research. *Journal of Trial and Error*

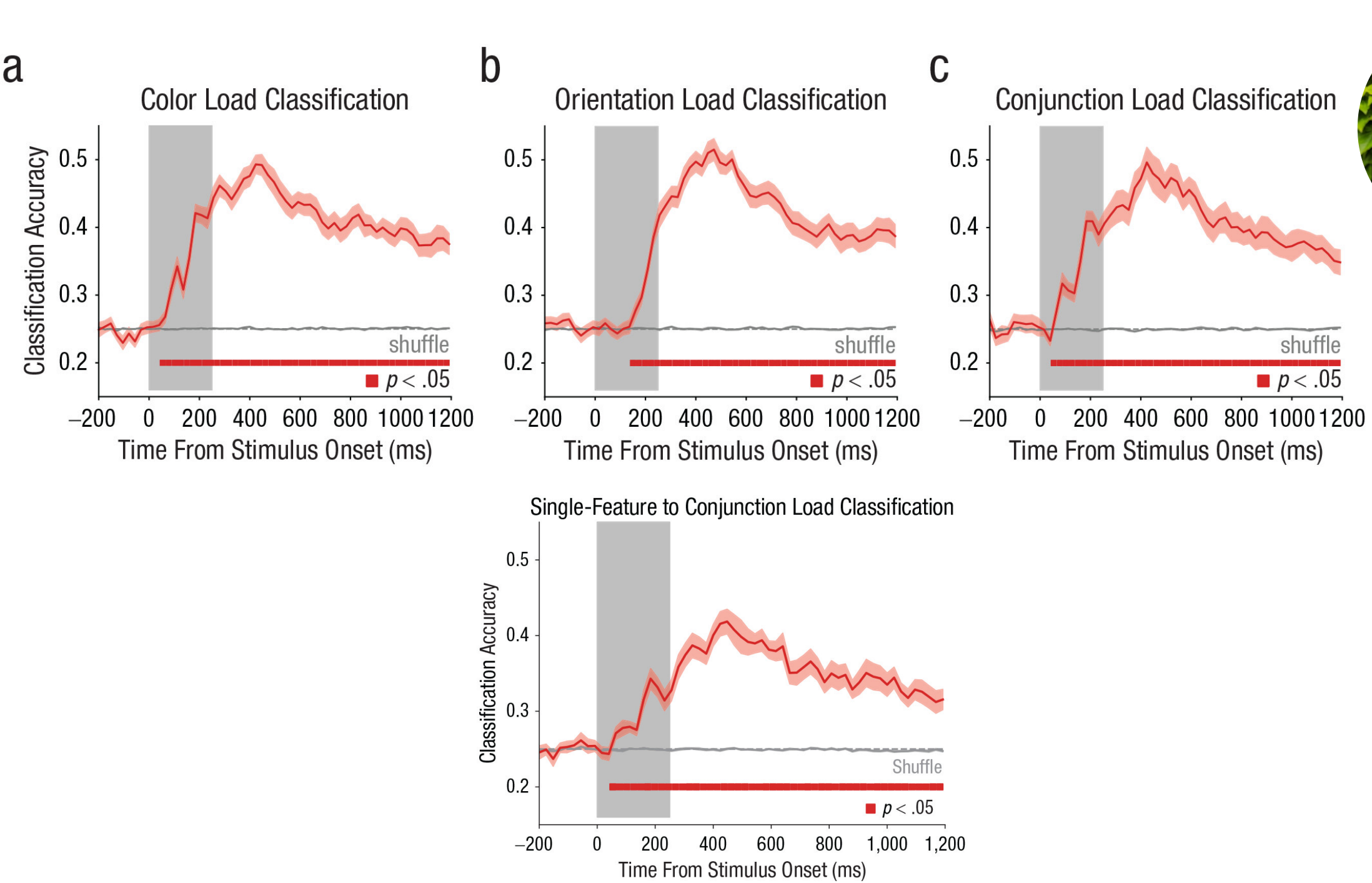
Personal benefits of Open Science

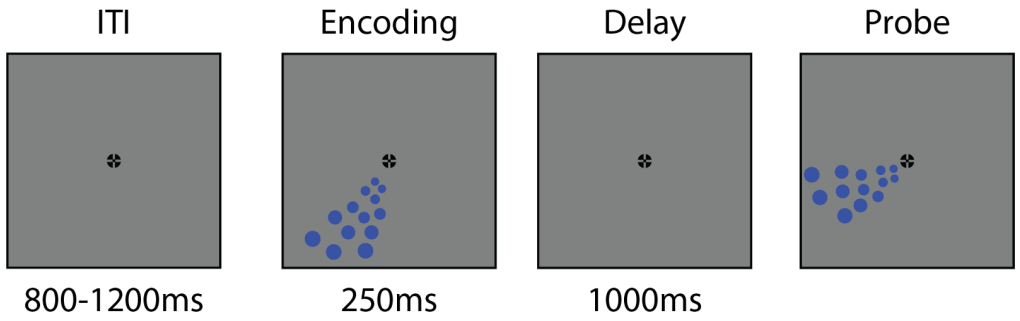
- May improve the quality and reliability of your scientific research
  - For example, preregistrations prompt theory development, justifications of sample sizes and analyses, and statistical power considerations to protect against researcher bias
- Increases the impact of your scientific research
  - Increase reviewers' quality of feedback if they reproduce your results and analyses
  - Increase citations from re-analysis and re-use of open datasets
- Can become part of your academic brand
  - Increasingly considered in grants and job applications

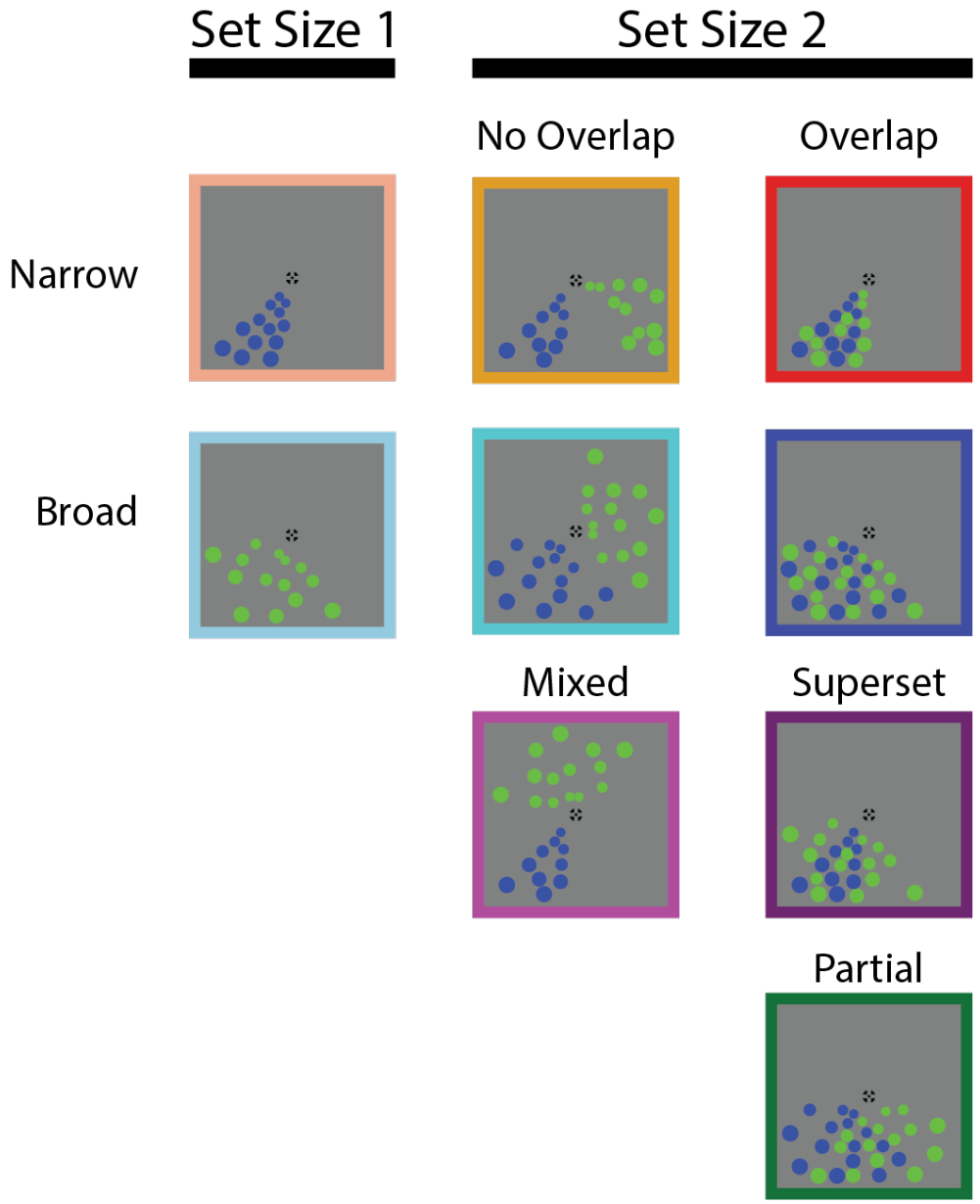
Markowetz, F. (2015). Five selfish reasons to work reproducibly. Genome biology, 16(1), 1-4.
Piwowar, H. A., & Vision, T. J. (2013). Data reuse and the open data citation advantage. *PeerJ*, *1*, e175.

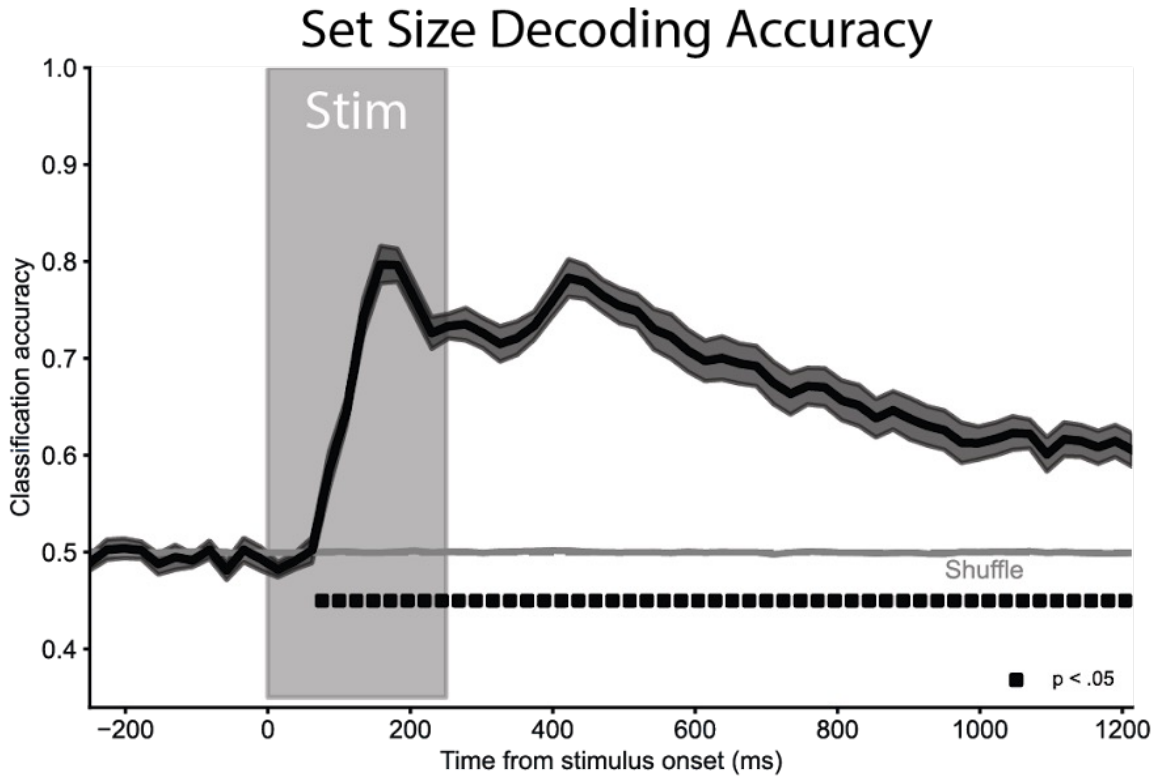
Hidden track 3:
What exactly are *pointers?*

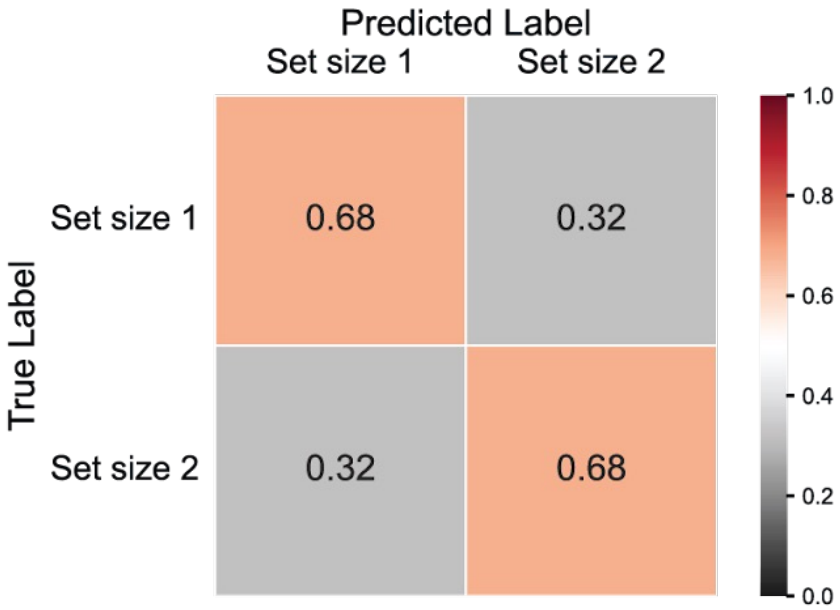

a
Experiment 1: Color
b
Experiment 2: Orientation
c
Experiment 3: Conjunction
600–1,000 ms
ITI
250 ms
Stimulus Array
1,000 ms
Retention
Until Response
Test Array


a
Color Load Classification
Classification Accuracy
0.5
0.4
0.3
0.2
shuffle
p < .05
–200 0 200 400 600 800 1000 1200
Time From Stimulus Onset (ms)
b
Orientation Load Classification
0.5
0.4
0.3
0.2
shuffle
p < .05
–200 0 200 400 600 800 1000 1200
Time From Stimulus Onset (ms)
c
Conjunction Load Classification
0.5
0.4
0.3
0.2
shuffle
p < .05
–200 0 200 400 600 800 1000 1200
Time From Stimulus Onset (ms)
Single-Feature to Conjunction Load Classification
Classification Accuracy
0.5
0.4
0.3
0.2
Shuffle
p < .05
–200 0 200 400 600 800 1,000 1,200
Time From Stimulus Onset (ms)

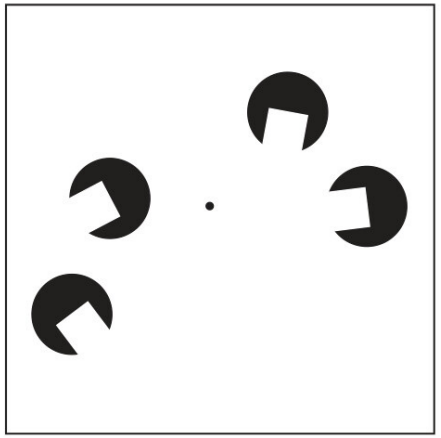

ITI
Encoding
Delay
Probe
800-1200ms
250ms
1000ms


Set Size 1
Set Size 2
No Overlap
Overlap
Narrow
Broad
Mixed
Superset
Partial

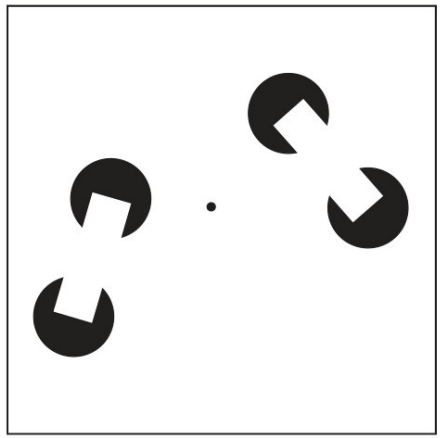

Set Size Decoding Accuracy
Stim
Classification accuracy
1.0
0.9
0.8
0.7
0.6
0.5
0.4
Shuffle
p < .05
−200
0
200
400
600
800
1000
1200
Time from stimulus onset (ms)

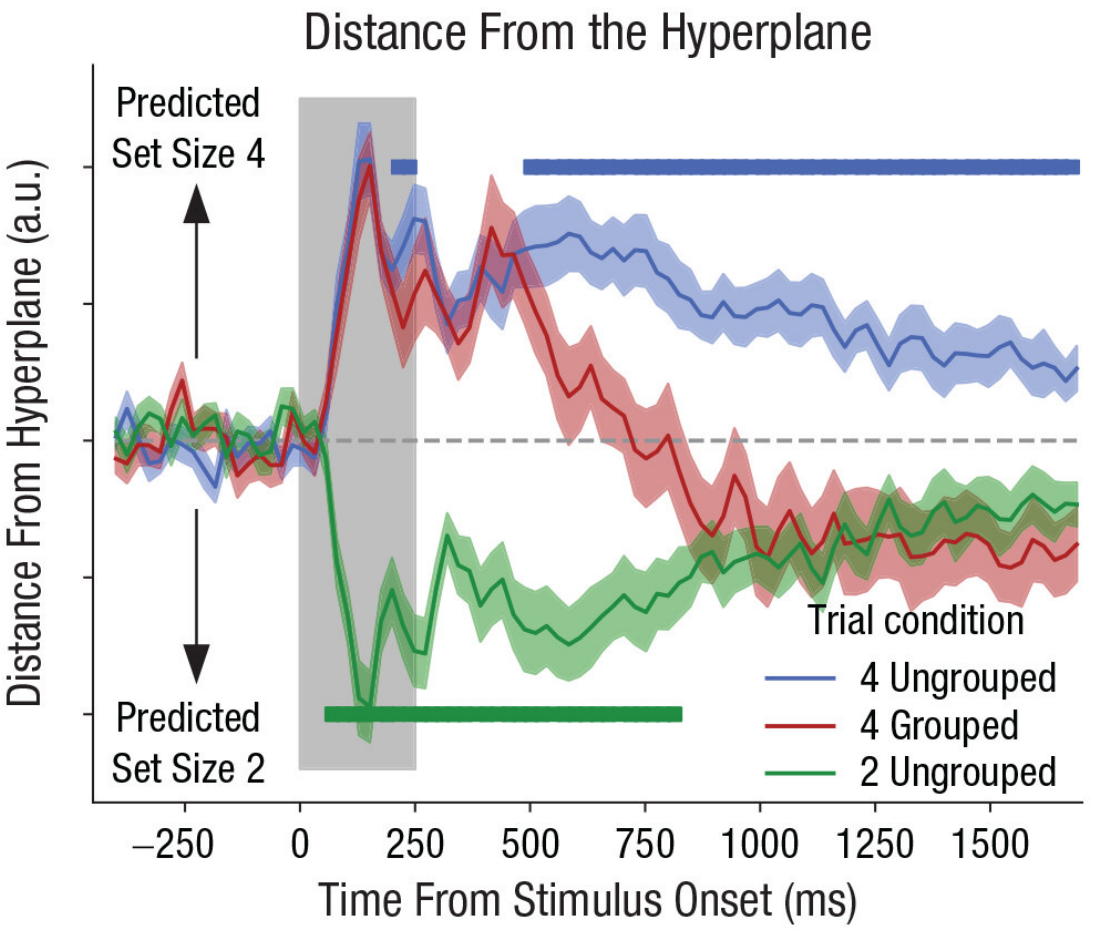

Predicted Label
Set size 1
Set size 2
True Label
Set size 1
Set size 2
0.68
0.32
0.32
0.68
1.0
0.8
0.6
0.4
0.2
0.0

Ungrouped

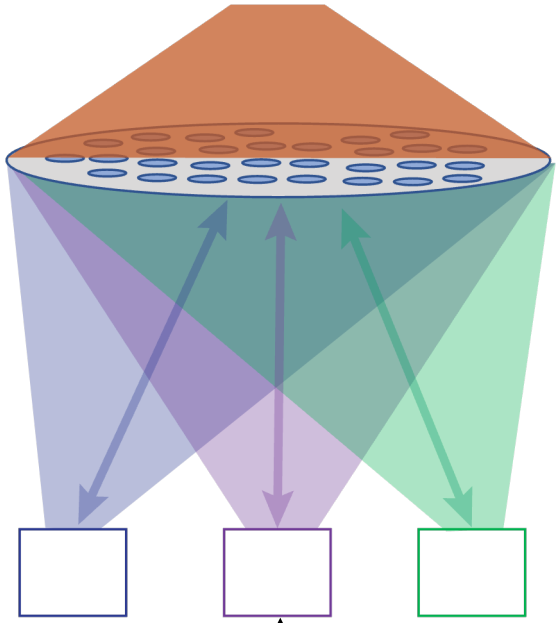


Grouped

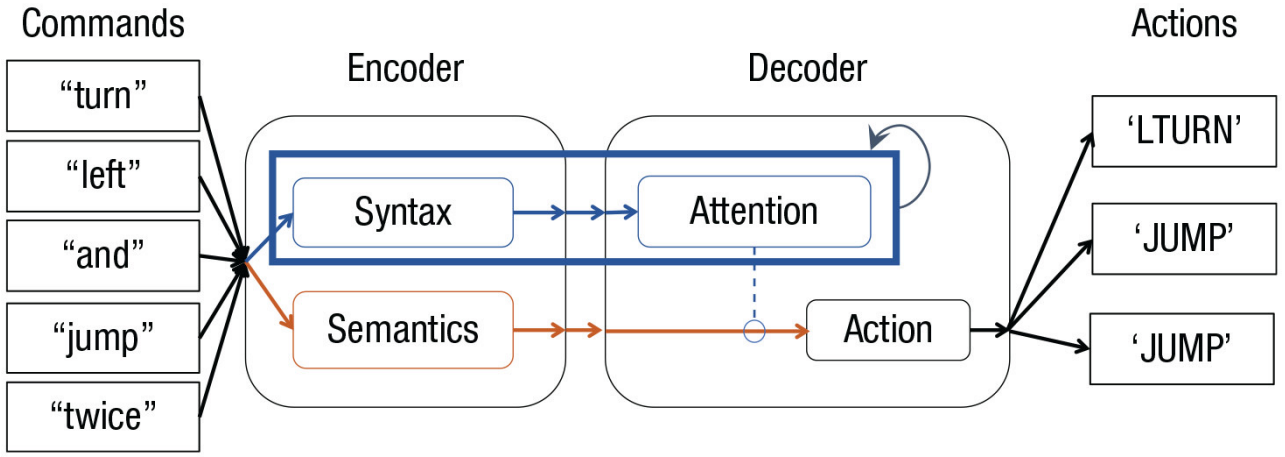


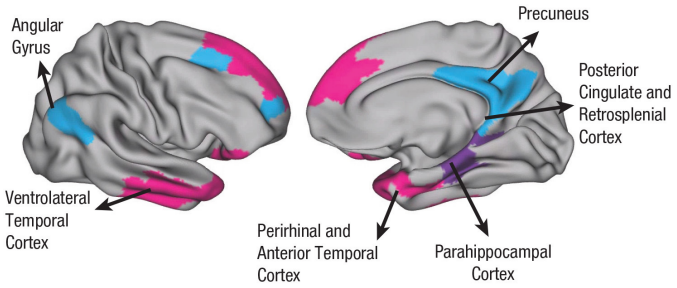


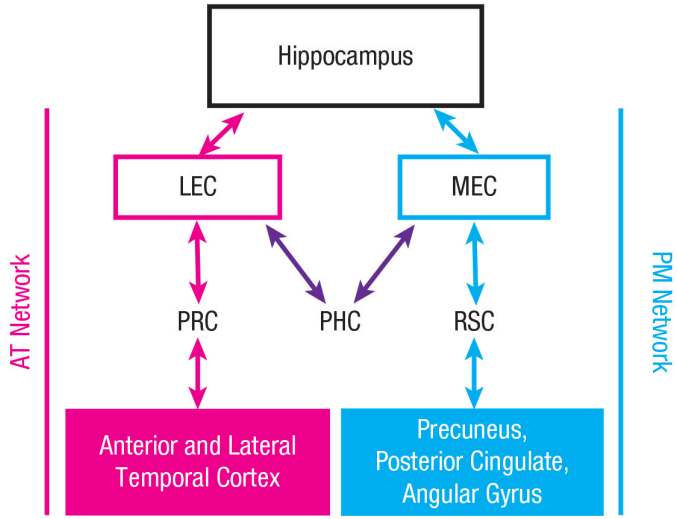
Hidden track 4:
What about *long-term memory?*



Connected to long-term memory

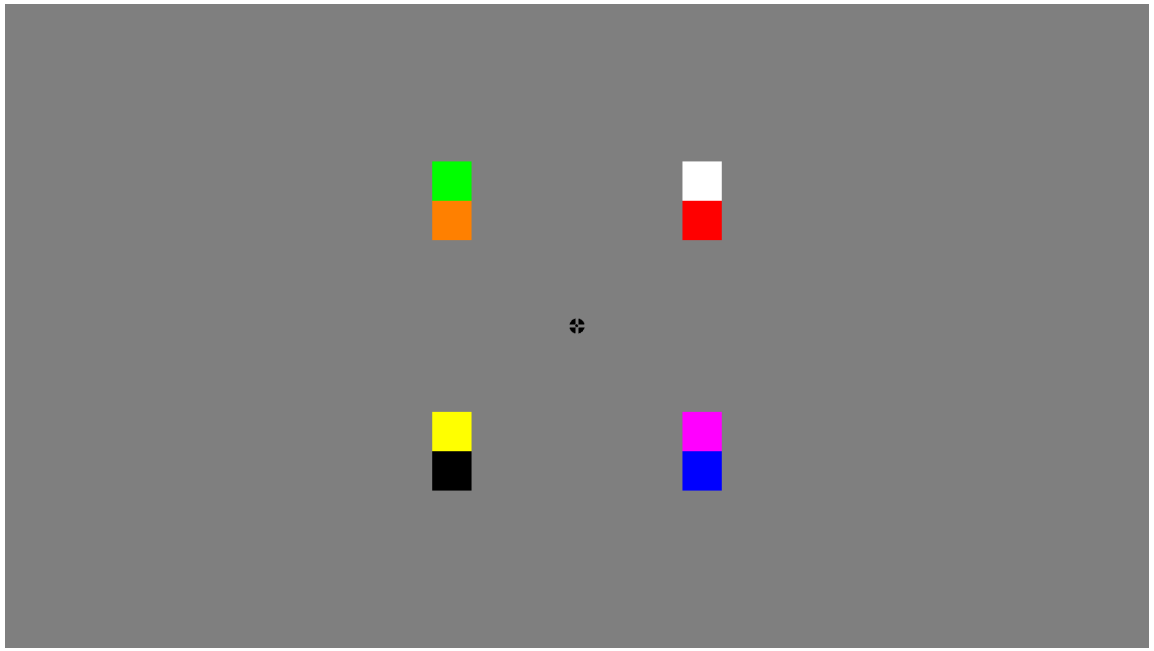


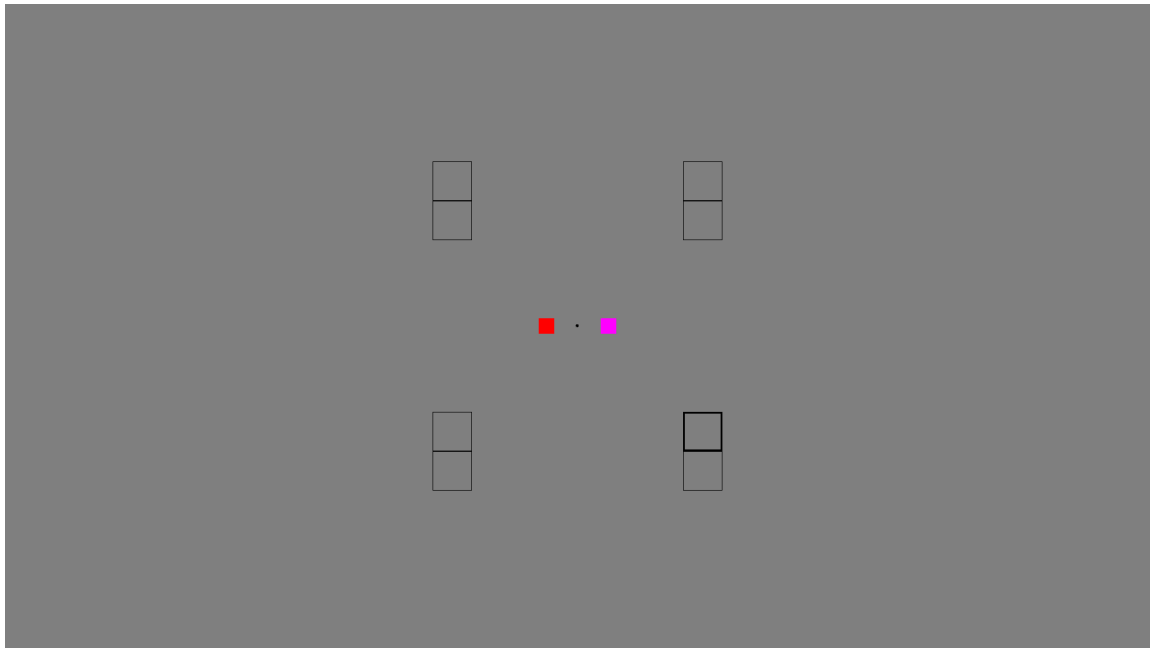




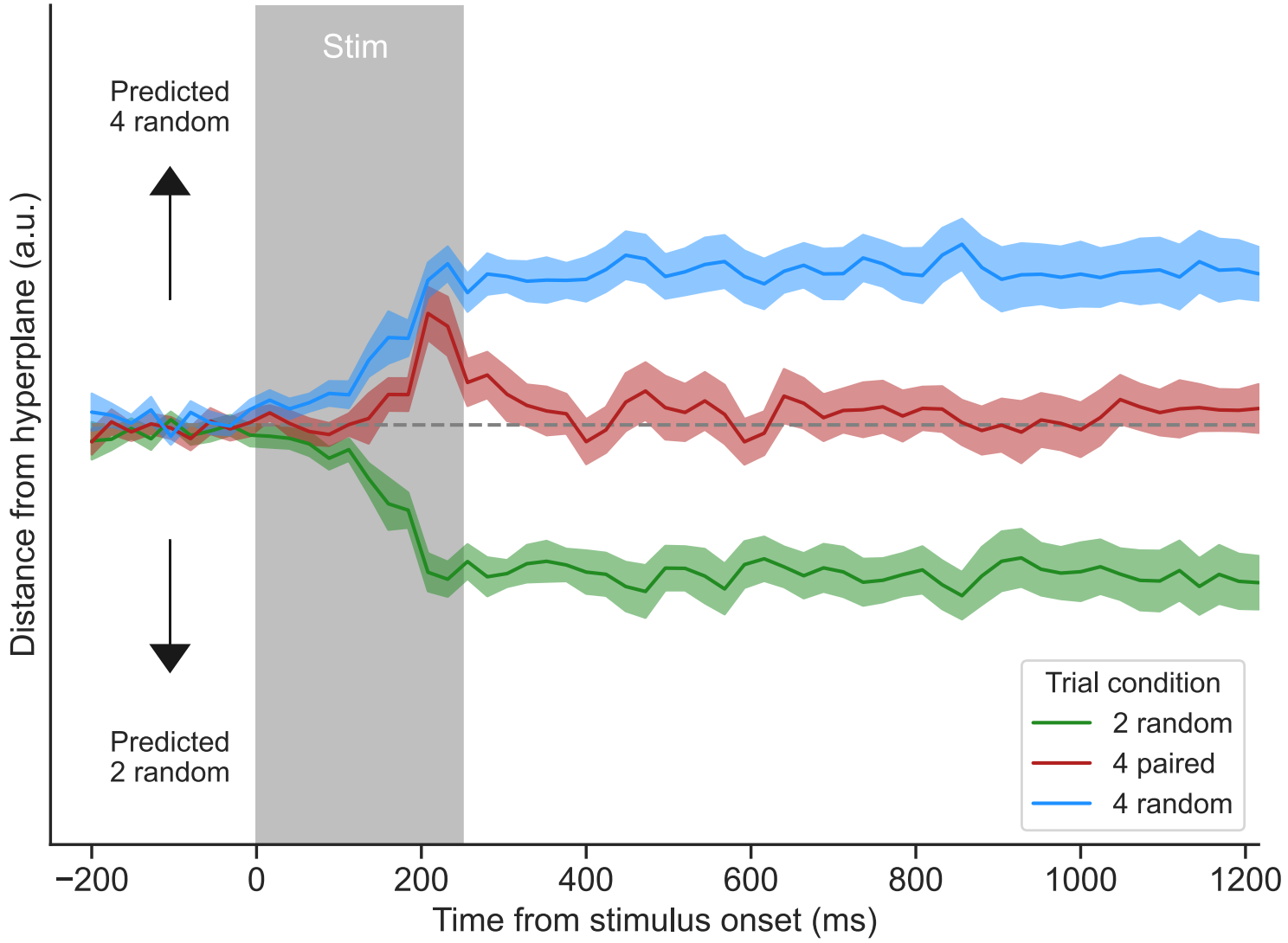
O'Reilly, R. C., Ranganath, C., & Russin, J. L. (2022). The structure of systematicity in the brain. *Current directions in psychological science*, *31*(2), 124-130.

Associative learning (“chunking”)

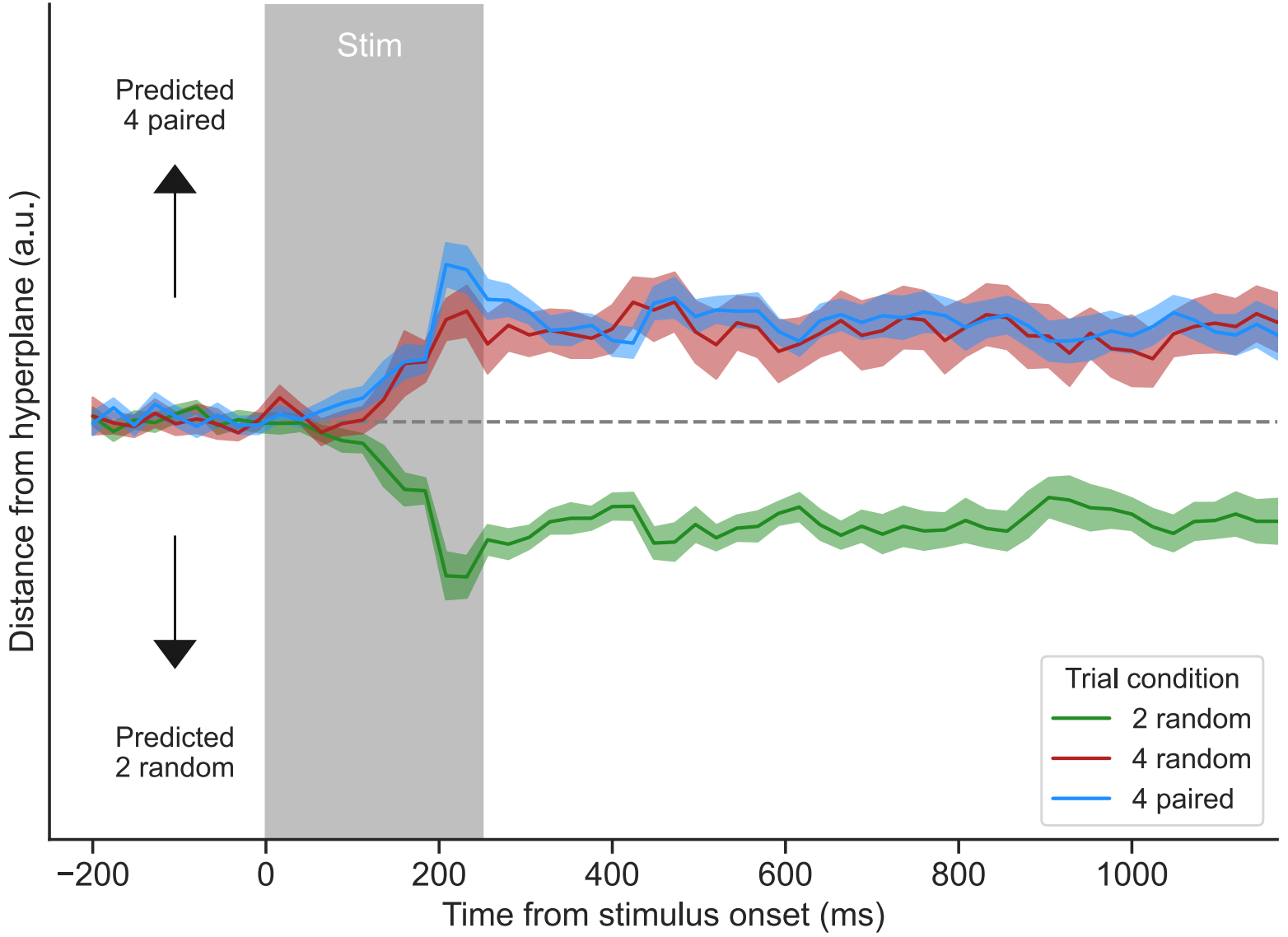




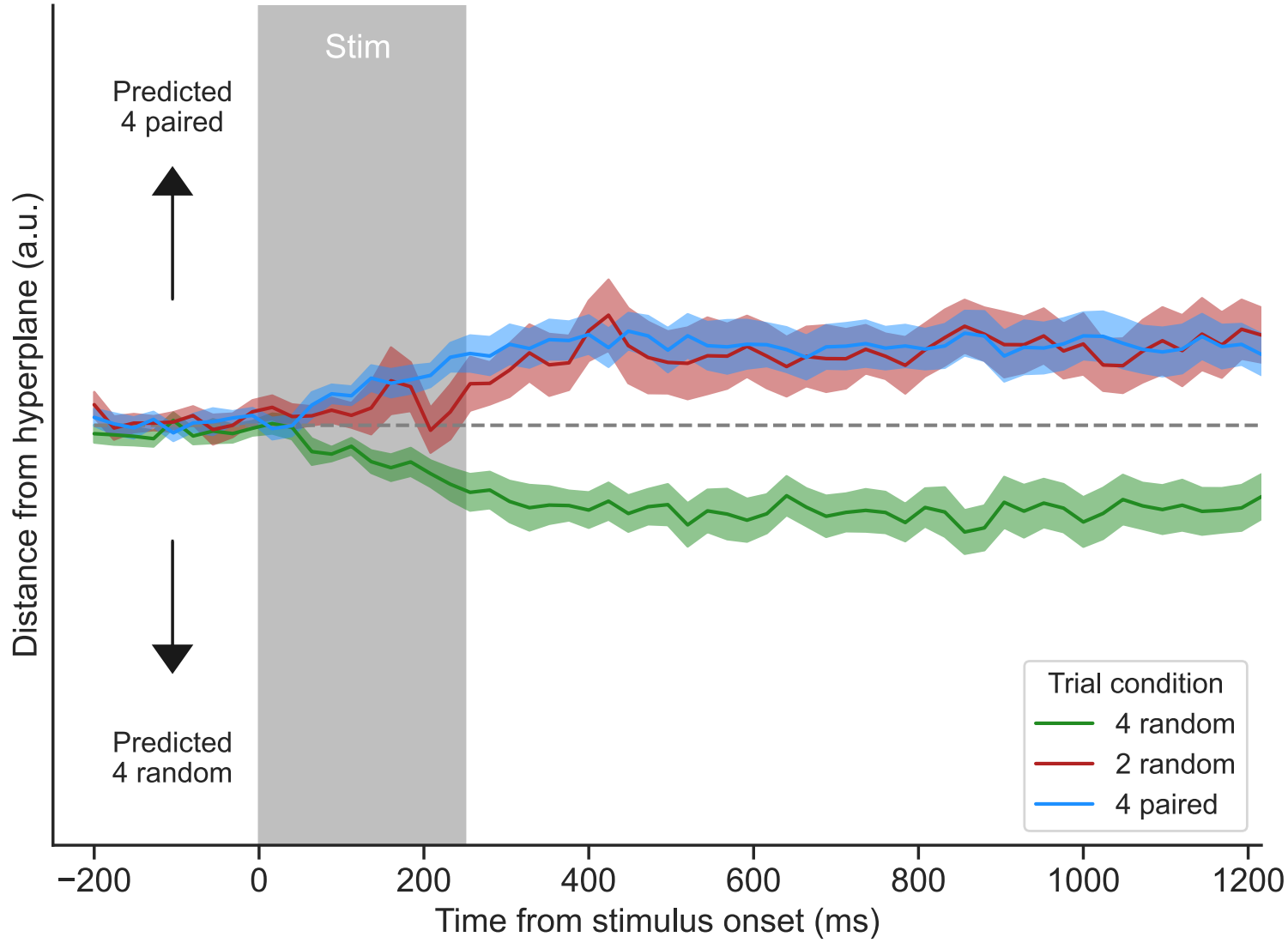
E2: 2 random versus 4 random



E2: 2 random versus 4 paired



E2: 4 random versus 4 paired



Multidimensional scaling