

# UC San Diego

## UC San Diego Electronic Theses and Dissertations

### Title

Inter-item interference and systematic biases in visual memory

### Permalink

<https://escholarship.org/uc/item/69t1n6gt>

### Author

Chunharas, Chaipat

### Publication Date

2019

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA SAN DIEGO

Inter-item interference and systematic biases in visual memory

A dissertation submitted in partial satisfaction  
of the requirements for the Doctor of Philosophy

in

Experimental Psychology

by

Chaipat Chunharas

Committee in charge:

Professor Vilayanur Subramanian Ramachandran, Co-Chair  
Professor John Thomas Serences, Co-Chair  
Professor Timothy Francis Brady  
Professor Eric Halgren  
Professor Viola Sophie Störmer  
Professor Edward Vul

2019

Copyright

Chaipat Chunharas, 2019

All Rights Reserved

The Dissertation of Chaipat Chunharas is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

---

---

---

---

---

Co-chair

---

Co-chair

University of California San Diego

2019

## DEDICATION

To my family, mentors, friends and science community.

## EPIGRAPH

Recall of memory is a creative process. What the brain stores is... only a core memory. Upon recall, this memory is then elaborated upon and reconstructed, with subtractions, additions, elaborations, and distortions.

Eric Kandel, *In search of memory*, 2006.

TABLE OF CONTENTS

Signature Page ..... iii

Dedication .....iv

Epigraph..... v

Table of Contents .....vi

List of Figures ..... vii

Acknowledgments..... viii

Vita.....xi

Abstract of the Dissertation ..... xii

Introduction..... 1

Chapter 1 Separating memoranda in depth increases  
visual working memory performance ..... 8

Chapter 2 Adaptive distortions in visual working memory .....26

Chapter 3 Selective amplification of salient features  
of visual memories during early memory consolidation.....52

## LIST OF FIGURES

Figure 1.1: The two target colors were presented in either the same or different depth planes in 3-D coordinates and either close or far in 2-D space.....	12
Figure 1.2: Results of Experiment 1 as a histogram of the responses centered around the target color, shown collapsed across all participants and conditions.....	13
Figure 1.3: Results of Experiment 1 in terms of the parameters from mixture modeling.....	14
Figure 1.4: Experimental procedure for Experiment 2.....	16
Figure 1.5: Main results of Experiment 2.....	17
Figure 1.6: The degree of positive correlation between depth-discrimination ability and performance on the visual-working-memory task.....	18
Figure 1.7: Participants who exhibited better depth-discrimination benefited more from the presence of depth information.....	19
Figure 2.1: Experimental procedure and results for Experiment 1.....	32
Figure 2.2: Experimental procedure and results for Experiment 2.....	34
Figure 2.3: Experimental procedure for Experiment 3.....	36
Figure 2.4: Experimental procedure for Experiment 4.....	37
Figure 2.5: Experimental procedure and results for Experiment 5.....	40
Figure 3.1: Data from Experiment 1.....	58
Figure 3.2: Timeline in Experiment 2.....	60
Figure 3.3: Results from Experiment 2 and 3.....	61
Figure 3.4: Results from Experiment 4.....	65



## ACKNOWLEDGEMENTS

I would like to thank Professor John Serences and Vilayanur Ramachandran for their support as co-chairs of my committee. Thank you for your great long support, guidance, and being my inspiration. It has been such an honor working under the supervision of great scientists like you two. Not everyone will want to make their bets on a random medical doctor who claims that he wants to be a good basic scientist. Thank you for the immeasurable faith you have in me. Thank you John for showing me how to approach problems in a careful and thoughtful manner. Thank you Rama for showing me that good scientific question can be fun and important at the same time.

Thank you, Dr. Eric Halgren. Without you, I will not have these amazing years of my life. Thank you for bringing me here and taking very good care of me and my family while I was under your supervision. I will try to be at least half as good as you as a supervisor and scientist.

I would like to thank Professor Timothy Brady for hundreds of hours of endless discussions. You open up so many doors of possibilities (I probably the only affiliated student who take your time more than your official students! And I'm proud of that for some reasons.). Thanks for all the fun times together. Thank you, Professor Viola Störmer, for allowing me to work with you in many projects. I learned a great deal observing how you formulate and tackle research question from both behavior and neural perspective. I will miss our Brady-Stoermer lab meetings a lot. Thank you, Dr. Adena Schachner, for a great opportunity to work under your supervision and expand my perspective on developmental psychology.

Thank you, Dr. Rosanne Rademaker, who is so many things to me. You are my friend, my mentor, my supporter, and many more. Thank you for believing in me (more than myself). I will not forget how we just casually take a short coffee walk and come up with many wonderful (and lot of time crazy) ideas. I cannot say enough how much I owe you (coffee and what's not). I will make sure I pay you back whenever and however I can. There are so many projects waiting for us in the future and I'm excited to continue working with you.

Thank you, my lab-mates from Perception and Cognition Lab, Center for Brain and Cognition, Vision and Memory lab, Multisensory integration and perception lab, Music and Developmental lab. Dr. Sirawaj Itthipuripat, Dr. Thomas Sprague, Dr. Mary Smith, Dr. Edward Ester. Dr. Kirsten Adam, Vy Vo, Dr. Stephanie Nelli, Nuttida Rungratsameetaweemana, Margeret Henderson, Anna Shafer-Skelton, Angus Chapman, Sunyoung Park, Timothy Sheehan, Miren Edelstein, Nicholas Root, Zeve Marcus, Jonas Lau, Jonathan Keefe, Jamal Williams, Frederik Geweke, Isabella Destefano, Hayden Schill, Minju Kim, Tanushree Agrawal. Thank you for being such wonderful researchers and friends. (And most of all, thanks for allowing me to pick your amazing brains).

My cohort - Lim Leong, Mingi Chung, Sasen Cain, Thijs Walbeek, Danbi Anh, Brendan Tomoschuk, Rob Louis, Kristin Donnelly, Isa Hutchings. Thanks for sharing our journey together, short or long. I will not be able to make it without your support.

I would also like to acknowledge all of my research assistants in alphabetical orders – Amanda Calcetas, Audrey Barszcz, Aylin Vega, Bridget Pak, Cameron Ryczek, Cara Spencer, Danielle Clarke, Hannah Heimer, Jacinda Duong, Justin Zhang, Kaushika Vayyala, Lara Hagopian, Melissa Vajanaphanich, Ming-ray Liao, Naomi Lee, Olga Chistokhodova, Omar Ahsan, Paloma Santos, Rachel Croft, Samson Peter, Shawn Anbiah, Wenjia Chen, and Youli Pan. It has been an amazing time working with all of you. This thesis will not exist without your great help. Thank you so much for your dedications.

Chapter 1, in full, is a reprint of the material as it appears in Journal of Vision, 2019, Chunharas, Chaipat; Rademaker, Rosanne; Sprague, Thomas; Brady, Timothy; Serences, John. The dissertation author was the primary investigator and author of this paper.

Chapter 2, in part, is currently being prepared for submission for publication of the material. Chunharas, Chaipat; Rademaker, Rosanne; Brady, Timothy; Serences, John. The dissertation author was the primary investigator and author of this material.

Chapter 3, in full, is a reprint of the material as it appears in PsyArxiv, 2018, Chunharas, Chaipat; Brady, Timothy; Ramachandran, Vilayanur. The dissertation author was the primary investigator and author of this paper.

Lastly, Thank you Bree and Emmalyn Chunharas, my family in Thailand and my Thai family here.

## VITA

- 2004            Doctor of Medicine, Chulalongkorn University, Thailand
- 2008            Diplomat of Internal Medicine, Khon Kaen University, Thailand
- 2011            Diplomat of Neurology, Chulalongkorn University, Thailand
- 2015            Master of science, University of California San Diego
- 2013-2019      Teaching Assistant, University of California San Diego
- 2013-2019      Research Assistant, University of California San Diego
- 2019            Doctor of Philosophy, University of California San Diego

## PUBLICATIONS

- Chunharas, C., Rademaker, R. L., Sprague, T. C., Brady, T. F., & Serences, J. T. (2019). Separating memoranda in depth increases visual working memory performance. *Journal of vision*, 19(1), 1-16.
- Chunharas, C., Rademaker, R. L., Brady, T. F., & Serences, J. T. (2019). Adaptive distortions in visual working memory. *PsyArXiv*.
- Chunharas, C., Brady, T. F., & Ramachandran, V. S. (2018). Selective amplification of salient features of visual memories during early memory consolidation. *PsyArXiv*.
- Rademaker, R. L., Chunharas, C., & Serences, J. T. (2018). Simultaneous representation of sensory and mnemonic information in human visual cortex. *bioRxiv*, 339200.
- Ramachandran, V. S., Chunharas, C., Croft, R., & Batal, N. (2018). Effect of Grouping, Segmentation, and Vestibular Stimulation on the Autokinetic Effect. *i-Perception*, 9(1), 2041669517752716.
- Ramachandran, V.S., Chunharas, C., Marcus, Z., Furnish, T., & Lin, A. (2018). Relief from intractable phantom pain by combining psilocybin and mirror visual-feedback (MVF). *Neurocase*, 24(2), 105-110.
- Ramachandran, V. S., Chunharas, C., & Smythies, M. (2017). Trigger Features for Conveying Facial Expressions: The Role of Segmentation. *i-Perception*, 8(6), 2041669517737792.
- Ramachandran, V. S., Vajanaphanich, M., & Chunharas, C. (2016). Calendars in the brain; their perceptual characteristics and possible neural substrate. *Neurocase*, 22(5), 461-465.
- Ramachandran, V. S., Chunharas, C., & Marcus, Z. (2016). Hypothesis concerning embodied calendars: A case study of number form, color spreading, and taste-color synaesthesia. *Medical Hypotheses*, 94, 58-62.
- Chunharas, C., & Ramachandran, V. S. (2016). Out of the Shadows. *Scientific American Mind*, 27(4), 56-61.

## ABSTRACT OF THE DISSERTATION

Inter-item interference and systematic biases in visual memory

by

Chaipat Chunharas

Doctor of Philosophy in Experimental Psychology

University of California San Diego, 2019

Professor Vilayanur S Ramachandran, Co-Chair  
Professor John T Serences, Co-Chair

Although our experience might convince us that our memory is limitless, researches showed that we could not precisely remember beyond 4-5 colors or orientations. One reason is that multiple items interfere and compete for limited resources. Interestingly, items not only interfere but also systematically “distort” each other in ways that make items more or less similar (i.e., attraction or repulsion biases). The main goal of the thesis is to investigate the complex relationships between multiple items in visual memory. We showed that inter-item interference could be reduced, hence memory performance is improved, by presenting stimuli further apart in 2D or 3D spatial position. Furthermore, we demonstrated inter-item systematic biases using a simple visual feature (color) as well as demonstrated the bias in more complex visual features (geometric shape and size). Lastly, we developed a general framework to explain

when and why we sometime see attraction or repulsion biases. We propose that attraction and repulsion biases reflect different goals of memory system – to summarize or reduce confusion respectively. We showed that the repulsion bias was stronger when we make two memory items more confusable (colors were very similar, presentation time was very short, memory delay was longer). Importantly, the repulsion is stronger in subjects with better general performance (measuring from an independent portion of data) – suggesting an adaptive nature rather than lack of effort to produce precise responses. Using the same paradigm, we showed that the direction was switched to attraction bias when we promoted summarizing than discriminating by increasing the memory load (from 2 to 4 items). Due to our biological limitation, our brains could not and do not try to remember the “truth.” Rather, our memory is only a distorted truth, and that does not matter as long as it is still useful.

## INTRODUCTION

When we are trying to remember multiple things, for example, twenty new faces, we tend to forget or misremember at least some of them. This is because our brain is not a camera or a video recorder. It cannot and does not try to represent the “truth” in great details. As a result, previous researches showed that memory becomes less precise when the memory system is pushed to the limit e.g., there are more things to remember or when subjects have to maintain memory for a long time. However, our subjective experience seems to tell us otherwise. We seem to effortlessly relive our pasts in great details – whether it just happened or happened a long time ago. On the other hand, research indicates that we cannot remember beyond 4-5 colored dots or  $7 \pm 2$  digits for longer than a few seconds<sup>1,2</sup>. This is due to limited resources that we have and, as a result, multiple items compete and interfere<sup>3-5</sup>.

How can the same brain that fails to hold more than few items after a short delay without interference holds so much memory or at least hold enough information for us to get through our days? The answer, perhaps, is that our memory system never encodes, maintains or retrieves information in isolation. On the contrary, it maximizes the information it can hold by constantly remembering new information in a face of any other available information it has. This can be a life-long past experience, an immediate past, or other things it is currently trying to remember. Together, they provide some frames of references that people can use to stabilize their memory.

Indeed, these reference frames have been revealed by investigating the systematic errors that people make when trying to recall information. These systematic biases in memory are well established in many domains (e.g., verbal or visual memory), at different levels of abstraction (color or orientation memory to face memory), and at different time courses (e.g., working memory or long-term memory)<sup>6-14</sup>. One classic example is for visual working memory of line orientations and colors. In orientation working memory tasks, people have to remember an orientation of a line that slightly tilted away from a perfectly vertical or horizontal (cardinal axes). When they later reproduce the remembered orientation after a short delay, the line was tilted away from cardinal axes even further (e.g.,  $5^\circ$  away from vertical axis becomes  $7^\circ$ ; repulsion bias)<sup>15-18</sup>. In color working memory tasks, people have to remember the color and later

click on a continuous color-wheel, trying to reproduce the exact hue of the remembered color. The reproduced color is systematically shifted toward the canonical color (e.g., the red-ish color becomes more typical red, indicating an attraction bias)<sup>19-21</sup>. However, the factors that give rise to repulsion biases in orientation memory and attraction biases in color memory are not totally clear. Nonetheless, it both demonstrates that memory is influenced by prior experiences (about cardinal axes and color category). In addition to the biases due to long-term priors as described above, memory is also influenced by recent events. For example, in standard orientation working memory tasks, subjects have to reproduce a remembered orientation after a short delay. The response on the current trial (say  $n$ th trial) is attracted towards the orientation from the previous trials ( $n-1$ th,  $n-2$ th, ... trials; the effect is weaker for more distant trials). This phenomenon is called serial dependence<sup>8,22,23</sup>.

The third class of systematic biases is that memory of an individual item is affected by other items that are concurrently held in memory. Going back to the first example, consider trying to remember twenty new faces in memory at the same time. One might compress the information about the twenty different faces into a summary representation such as “there are about this many faces, and this is how the mean face looks in general”. Indeed, when one particular face in the group is later probed, the memory of that face is distorted toward the group average<sup>9,24</sup>. This kind of attraction bias has been shown with other types of stimuli as well e.g. size, color, etc.<sup>25,26</sup>. The function is to stabilize the behavioral responses. As the errors of individual items are systematically leaning toward the mean, the overall errors are reduced. Meanwhile, repulsion biases have also been shown in other studies<sup>27,28</sup>. To date, there is still no clear theoretical framework that captures why and when repulsion vs. attraction biases will occur.

My thesis aims to understand the mechanisms of memory which allow us to remember multiple items by studying how representations of items in memory interact. In chapter 1, I revisited the idea that a major bottleneck in memory is due to inter-item interference caused by shared resources. Previous research has shown that memory is more precise when multiple items are presented farther away from each other (as opposed to closer together) and thus items may compete more in given the map-like organization of neural resources. I pushed this idea further by asking if separating multiple items in depth



would have similar or different effects as memory for spatial position. To test these ideas, I created a virtual three-dimensional display using stereo-goggle and binocular depth cues. In the first experiment, I asked people to remember two color dots and to later report the exact hue of one of them (i.e., target memory color) by choosing from a color wheel. Importantly, two memory items were either presented closer or farther away in 2D spatial position as well as same or different depth plane (3D). I found that the responses were more precise when two items were either farther apart in 2D or 3D space, and there was no interaction between these factors – suggesting that the interference does not exclusively happen at retinotopic level but also happen later stage where second-order 3D perception was constructed and perceived. In experiment 2, I further ask if the benefit was even more pronounced when there were more items to remember. To this end, I manipulated the display such that there are 2, 4, 6, 8, or 12 to-be-remembered items that were either presented on the same or different depth planes. I found separation in depth help improves memory performance in subjects who can clearly see the depth in our display. I hypothesize that the benefits due to separation in depth might help people from being overwhelm by amount of information (esp. when there are many items to remember) and be able. to deploy their attention more effectively (i.e., attentional filtering mechanisms <sup>29,30</sup>). Additionally, I consistently found that responses to the target memory color were systematically repelled from non-target memory colors, replicating previous findings <sup>27</sup>. Interestingly, there were no significant differences between conditions – suggesting that the repulsion bias is not merely a well-known color contrast at a perceptual level which should be stronger when two colors are closer in space.

In chapter 2, I did a more detailed investigation about when and why repulsion bias occurs. Specifically, I hypothesized that repulsion biases reflect a strategic choice to keep our memories distinct. Going back to the face example, consider a scenario that involves not remembering all twenty faces but instead trying to remember two important people while not confusing them. Since the goal is now to avoid confusion, the attempt to compress the information to compute a summary representation would hurt behavioral performance. Instead, amplifying differences between the two faces, thus inducing a repulsion bias, would be more optimal. Under this hypothesis, one would expect even stronger repulsion

biases if those two persons were extremely difficult to discriminate (e.g. identical twins) (of course, unless the differences are unperceivable). This also includes other metrics that might increase confusion such as weaker signals or weaker memory strength. In a series of experiments, I showed that repulsion biases are stronger when two colors were more similar. And when I gave people more time to remember (hence, they had stronger memory signals), the colors needed to be even more similar to yield similar repulsion biases. Moreover, the repulsion is also stronger when the delay time between seeing a display and testing was longer. Next, I asked if repulsion actually reflects a lack of effort to remember or produce accurate memories by measuring general task performance. Contrary to this idea, I found that there was stronger repulsion bias in subjects with better performance, suggesting that the repulsion bias might be an adaptive process. Lastly, I added more items to remember so that summarizing is more appropriate than discriminating. As expected, I found attraction biases instead of repulsion. Together, these experiments create a coherent framework that helps reconcile seemingly disparate findings.

In the last chapter, to demonstrate that this framework can be generalized to other domains and timescales and response modes, I investigated systematic biases in long-term memory using higher order visual features like geometric shapes and sizes via a free-drawing task. I found that when people have to remember a rectangle and a square after many hours, the rectangle became longer than the original rectangle. Similar repulsion biases also revealed that when people have to remember two squares with different sizes, meaning that the size difference between two squares is larger after a long delay period.

In three chapters, my thesis demonstrates how multiple items interact in visual memory. I showed how they interfere and how the interference can be potentially reduced by separating items in depth or 2D spatial positions. I also showed that multiple items do not merely compete for shared resources but also distort each other in systematic and adaptive manners. I clearly show that we do not remember each individual item separately but use how multiple items relate to one another to constraint noises and maintain discriminability more efficiently. The directions of the biases depend on whether the current goal is to summarize or discriminate, resulting in attraction and repulsion biases respectively.

On a surface, memory distortions can be seen as undesirable outcomes. But instead of a failure of the memory system to remain truthful, it can also be seen as an intelligent strategy to overcome its biological limitations and allow us to achieve our goals. With this perspective in mind, it is conceivable that some types of memory deficit are due to some forms of dysfunctions in these mechanisms – either an inability to make use of overall inter-item relationships or imbalance between generalization vs. discrimination (too much attraction or repulsion bias which lead to failure to discriminate or failure to generalize from the memory respectively).

In fact, the memory system is not the only system that make use of contexts or reference frames. For example, our perceptions are only possible when infinite possibilities are reduced and constraint by our knowledge or context. Since our perception is already an inferential process (non-veridical), it is rather odd to ever hope for veridical memories. On contrary, one can think of memory representations as the noisier version of perceptual representations – as they are no longer anchored to the incoming inputs. As the signal is getting weaker, the representations should rely even more on the context. While it shows us how limited our memory system is, it also shows us how efficient our system can be. Our brains could not and do not try to remember the “truth.” Rather, our memory is only a distorted truth, and that does not matter as long as it is still useful.

## References

1. Miller, G. A. The magical number seven plus or minus two: some limits on our capacity for processing information. *Psychol. Rev.* **63**, 81–97 (1956).
2. Luck, S. J. & Vogel, E. K. The capacity of visual working memory for features and conjunctions. *Nature* **390**, 279–281 (1997).
3. Bays, P. M. Spikes not slots: noise in neural populations limits working memory. *Trends Cogn. Sci.* **19**, 431–438 (2015).
4. Ma, W. J., Husain, M. & Bays, P. M. Changing concepts of working memory. *Nat. Neurosci.* **17**, 347–356 (2014).
5. Van den Berg, R., Shin, H., Chou, W.-C., George, R. & Ma, W. J. Variability in encoding precision accounts for visual short-term memory limitations. *Proceedings of the National Academy of Sciences*

- 109**, 8780–8785 (2012).
6. Brady, T. F., Konkle, T. & Alvarez, G. A. A review of visual memory capacity: Beyond individual items and toward structured representations. *J. Vis.* **11**, 4 (2011).
  7. Alvarez, G. A. Representing multiple objects as an ensemble enhances visual cognition. *Trends Cogn. Sci.* **15**, 122–131 (2011).
  8. Fischer, J. & Whitney, D. Serial dependence in visual perception. *Nat. Neurosci.* **17**, 738–743 (2014).
  9. Liberman, A., Fischer, J. & Whitney, D. Serial dependence in the perception of faces. *Curr. Biol.* **24**, 2569–2574 (2014).
  10. Huang, J. & Sekuler, R. Distortions in recall from visual memory: two classes of attractors at work. *J. Vis.* **10**, 24.1–27 (2010).
  11. Rademaker, R. L., Bloem, I. M., De Weerd, P. & Sack, A. T. The impact of interference on short-term memory for visual orientation. *J. Exp. Psychol. Hum. Percept. Perform.* **41**, 1650–1665 (2015).
  12. Bartlett, F. C. Some Experiments on the Reproduction of Folk-Stories. *Folklore* **31**, 30–47 (1932).
  13. Loftus, E. F. Planting misinformation in the human mind: a 30-year investigation of the malleability of memory. *Learn. Mem.* **12**, 361–366 (2005).
  14. Roediger, H. L., & McDermott, K. B. Creating false memories: Remembering words not presented in Lists. *J. Exp. Psychol. Learn. Mem. Cogn.* **21**, 803 (1995).
  15. Girshick, A. R., Landy, M. S. & Simoncelli, E. P. Cardinal rules: visual orientation perception reflects knowledge of environmental statistics. *Nat. Neurosci.* **14**, 926–932 (2011).
  16. Wei, X.-X. & Stocker, A. A. A Bayesian observer model constrained by efficient coding can explain 'anti-Bayesian' percepts. *Nat. Neurosci.* **18**, 1509 (2015).
  17. Jastrow, J. Studies from the University of Wisconsin: on the judgment of angles and positions of lines. *Am. J. Psychol.* **5**, 214–248 (1892).
  18. Smith, S. L. Angular estimation. *J. Appl. Psychol.* **46**, 240–246 (1962).
  19. Bae, G.-Y., Olkkonen, M., Allred, S. R. & Flombaum, J. I. Why some colors appear more memorable than others: A model combining categories and particulars in color working memory. *J. Exp. Psychol. Gen.* **144**, 744–763 (2015).
  20. Boynton, R. M., Fargo, L., Olson, C. X. & Smallman, H. S. Category effects in color memory. *Color Res. Appl.* **14**, 229–234 (1989).
  21. Kelly, L. J. & Heit, E. Recognition memory for hue: Prototypical bias and the role of labeling. *J. Exp. Psychol. Learn. Mem. Cogn.* **43**, 955–971 (2017).
  22. Bliss, D. P., Sun, J. J. & D'Esposito, M. Serial dependence is absent at the time of perception but increases in visual working memory. *Sci. Rep.* **7**, 14739 (2017).

23. Kiyonaga, A., Scimeca, J. M., Bliss, D. P. & Whitney, D. Serial Dependence across Perception, Attention, and Memory. *Trends Cogn. Sci.* **21**, 493–497 (2017).
24. Haberman, J. & Whitney, D. Seeing the mean: ensemble coding for sets of faces. *J. Exp. Psychol. Hum. Percept. Perform.* **35**, 718–734 (2009).
25. Brady, T. F. & Alvarez, G. A. Hierarchical encoding in visual working memory: ensemble statistics bias memory for individual items. *Psychol. Sci.* **22**, 384–392 (2011).
26. Brady, T. F. & Alvarez, G. A. Contextual effects in visual working memory reveal hierarchically structured memory representations. *J. Vis.* **15**, 6 (2015).
27. Golomb, J. D. Divided spatial attention and feature-mixing errors. *Atten. Percept. Psychophys.* **77**, 2562–2569 (2015).
28. Bae, G. Y., & Luck, S. J. Interactions between visual working memory representations. *Atten. Percept. Psychophys.* **79**, 2376–2395 (2017).
29. Cowan, N. & Morey, C. C. Visual working memory depends on attentional filtering. *Trends Cogn. Sci.* **10**, 139–141 (2006).
30. Fukuda, K. & Vogel, E. Attentional filtering efficiency and individual differences in VSTM capacity. *J. Vis.* **7**, 684–684 (2010).

## Chapter 1

Separating memoranda in depth increases visual working  
memory performance

# Separating memoranda in depth increases visual working memory performance

**Chaipat Chunharas**

Psychology Department,  
University of California San Diego, La Jolla, CA, USA  
King Chulalongkorn Memorial Hospital,  
Chulalongkorn University, Bangkok, Thailand



**Rosanne L. Rademaker**

Psychology Department,  
University of California San Diego, La Jolla, CA, USA  
Donders Institute for Brain, Cognition and Behavior,  
Radboud University, Nijmegen, the Netherlands



**Thomas C. Sprague**

Department of Psychology,  
University of California Santa Barbara,  
Santa Barbara, CA, USA



**Timothy F. Brady**

Psychology Department,  
University of California San Diego, La Jolla, CA, USA



**John T. Serences**

Psychology Department,  
University of California San Diego, La Jolla, CA, USA  
Neurosciences Graduate Program,  
University of California San Diego, La Jolla, CA, USA  
Kavli Institute for Brain and Mind,  
University of California, San Diego, La Jolla, CA, USA



Visual working memory is the mechanism supporting the continued maintenance of information after sensory inputs are removed. Although the capacity of visual working memory is limited, memoranda that are spaced farther apart on a 2-D display are easier to remember, potentially because neural representations are more distinct within retinotopically organized areas of visual cortex during memory encoding, maintenance, or retrieval. The impact on memory of spatial separability in depth is less clear, even though depth information is essential to guiding interactions with objects in the environment. On one account, separating memoranda in depth may facilitate performance if interference between items is reduced. However, depth information must be inferred indirectly from the 2-D retinal image, and less is known about how visual cortex represents depth. Thus, an alternative possibility is that separation in depth does not attenuate between-items interference; it may even impair performance, as attention must be distributed across a larger volume of 3-D space. We tested these alternatives using a stereo display while

participants remembered the colors of stimuli presented either near or far in the 2-D plane or in depth. Increasing separation in-plane and in depth both enhanced performance. Furthermore, participants who were better able to utilize stereo depth cues showed larger benefits when memoranda were separated in depth, particularly for large memory arrays. The observation that spatial separation in the inferred 3-D structure of the environment improves memory performance, as is the case in 2-D environments, suggests that separating memoranda in depth might reduce neural competition by utilizing cortically separable resources.

## Introduction

Visual working memory (VWM) supports the integration of past and present sensory information via short-term maintenance when such information is no longer directly accessible. Performance on VWM tasks

Citation: Chunharas, C., Rademaker, R. L., Sprague, T. C., Brady, T. F., & Serences, J. T. (2019). Separating memoranda in depth increases visual working memory performance. *Journal of Vision*, 19(1):4, 1–16, <https://doi.org/10.1167/19.1.4>.

<https://doi.org/10.1167/19.1.4>

Received July 8, 2018; published January 11, 2019

ISSN 1534-7362 Copyright 2019 The Authors

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.



is highly correlated with measures of general intelligence and other related outcome measures and is therefore thought to reflect a core cognitive capacity (Baddeley, 1986; Conway, Cowan, Bunting, Theriault, & Minkoff, 2002; Engle, Tuholski, Laughlin, & Conway, 1999; Fukuda, Vogel, Mayr, & Awh, 2010). In most VWM studies, simple visual stimuli are presented on a 2-D computer screen and participants remember specific features, such as color or orientation, that are presented at different spatial locations (Engle et al., 1999; Luck & Vogel, 1997; Simons & Levin, 1997; Zhang & Luck, 2008). Based on such work, VWM is known to be limited in capacity (Bays, Catalao, & Husain, 2009; Bays & Husain, 2008; Ma, Husain, & Bays, 2014; Schurgin, Wixted, & Brady, 2018), such that increasing the number of items to be remembered or the delay duration leads to reductions in memory precision (Ma et al., 2014; Panichello, DePasquale, Pillow, & Buschman, 2018; Rademaker, Park, & Sack, 2018; Shin, Zou, & Ma, 2017; van den Berg, Shin, Chou, George, & Ma, 2012; Zhang & Luck, 2008), reductions in confidence (Rademaker, Tredway, & Tong, 2012), the misbinding or “swapping” of different visual features (Bays, 2016; Bays, Gorgoraptis, Wee, Marshall, & Husain, 2011; Bays, Wu, & Husain, 2011), and the tendency to chunk information into group-level ensemble representations (Brady & Alvarez, 2011).

One of the key factors that govern interactions between remembered items is the degree to which different memoranda can be bound to distinct spatial locations. For example, detecting a change in a remembered object is more challenging when the spatial configuration of the display is modified between encoding and test, highlighting the importance of spatial layout and spatial location in VWM (Hollingworth, 2007; Hollingworth & Rasmussen, 2010; Jiang, Olson, & Chun, 2000; Olson & Marshuetz, 2005; Phillips, 1974; Postle, Awh, Serences, Sutterer, & D’Esposito, 2013; Treisman & Zhang, 2006). Memory performance is improved when multiple simultaneous memoranda are presented far from each other, compared to close to each other, suggesting a role for spatial interference (Cohen, Rhee, & Alvarez, 2016; Emrich & Ferber, 2012). Furthermore, presenting memoranda sequentially in different spatial locations leads to better memory performance compared to sequentially presenting items in the same spatial location, even when location is task irrelevant (Pertzov & Husain, 2014).

The importance of 2-D space in VWM is consistent with the clear maplike organization of 2-D spatial position across the cortical surface, which should result in less neural competition and more distinct representations as items are spaced farther apart (Engel, Glover, & Wandell, 1997; Grill-Spector & Malach, 2004; Maunsell & Newsome, 1987; Sereno et al., 1995;

Sereno, Pitzalis, & Martinez, 2001; Talbot & Marshall, 1941). This general idea is consistent with a sensory-recruitment account, which proposes that early sensory cortex supports the maintenance of sensory information in working memory (D’Esposito & Postle, 2015; Emrich, Riggall, Larocque, & Postle, 2013; Harrison & Tong, 2009; Pasternak & Greenlee, 2005; Rademaker, Chunharas, & Serences, 2018; Serences, 2016; Serences, Ester, Vogel, & Awh, 2009; Sreenivasan, Curtis, & D’Esposito, 2014). Thus, overlap or competition between representations in retinotopic maps may impose limits on how well visual information is encoded and remembered (Emrich et al., 2013; Sprague, Ester, & Serences, 2014).

The impact of presenting memoranda in different depth planes is less clear. Given that the retina encodes a 2-D projection of light coming from a complex 3-D environment, depth information must be indirectly inferred based on binocular cues like retinal disparity and monocular cues from pictorial depth indicators. In addition to the second-order nature of depth computations, there is also far less evidence of maplike 3-D spatial representations in visual cortex. However, a recent study suggests that there are topographic representations of depth encoded in some visual areas, so separation in 3-D may operate much like separation in 2-D (Finlayson, Zhang, & Golomb, 2017). In addition, studies of visual search suggest that 3-D structure may generally facilitate information processing. For example, visual-search performance is better when depth information is present, particularly when the 3-D structure of the display is kept constant across trials (McCarley & He, 2001). Visual-search performance is also substantially better when participants are searching for a combination of color and depth or motion and depth compared to searching for a combination of two visual features that are not separated in depth. This finding suggests that depth separation can facilitate the separate encoding of visual features (Nakayama & Silverman, 1986).

That said, the few previous studies that have directly investigated the effect of depth on VWM task performance have reported conflicting evidence, with some finding performance improvements and some finding performance decrements (Qian, Li, Wang, Liu, & Lei, 2017; Reeves & Lei, 2014; Xu & Nakayama, 2007). In addition, studies focusing on different aspects of information processing, such as selective attention, suggest that separating visual stimuli in depth might lead to impaired performance because encoding across different depth planes increases the total volume of 3-D space that participants must attentively monitor (Andersen, 1990; Andersen & Kramer, 1993; Atchley, Kramer, Andersen, & Theeuwes, 1997; Downing & Pinker, 1985; Enns & Rensink, 1990; Finlayson & Grove, 2015; Finlayson, Remington, Retell, & Grove,



2013; Theeuwes, Atchley, & Kramer, 1998). For instance, while attention tends to naturally spread across perceived 3-D surfaces, it is not as easy to divide attention between two 3-D surfaces (He & Nakayama, 1995). Similarly, separating memoranda in depth might hinder performance because of these limitations in attention. Thus, it remains unclear whether depth would be important in the same way as 2-D space for improving the separability of representations in working memory.

To test these alternative accounts, we examined the effects of 2-D in-plane and 3-D depth separation on memory precision (Experiment 1) and interactions between separation in depth and the number of remembered items (i.e., the set size of the memory array; Experiment 2). In Experiment 1, we found that separating items in depth improves memory performance in a manner similar to separating items in the 2-D plane. In Experiment 2, we found that the benefits of separating memoranda in depth were particularly evident in participants who were better able to perceive items in depth, and when participants had to remember a larger number of items. Together, these findings show that both 2-D in-plane and 3-D across-planes spatial separability improve VWM performance. Thus, performance benefits for items separated in the 2-D plane may extend to structured representations of the inferred 3-D layout of a visual scene, perhaps as a result of the recruitment of more retinotopically distinct neural resources.

## Experiment 1

### Methods

#### Participants

Thirty healthy volunteers (21 women, nine men; mean age [ $\pm$  standard error of the mean] = 20.87  $\pm$  0.53 years) from the University of California San Diego (UCSD) community participated in the experiment. All procedures were approved by the UCSD Institutional Research Board. All participants reported normal or corrected-to-normal vision without color-vision deficiency, and provided written informed consent. To ensure that all participants had stereovision, we prescreened for stereo blindness by asking all participants to look at a random-dot stereogram display through binocular goggles and then identify three different geometric shapes (a triangle, a square, and a circle). These shapes can be seen only if participants successfully fuse the images from the left and right eyes. All participants in this study correctly identified all three shapes. Participants were unaware of the purpose of the study and received course credit for their time.

Three participants were excluded from the analysis due to low performance (circular standard deviation of more than 45°).

#### Stimuli and procedure

Stimuli were rendered using virtual-reality goggles (Oculus DK2, Microsoft, Redmond, WA) with a resolution of 1,920  $\times$  1,080, at a 60-Hz refresh rate and a screen size of 12.6  $\times$  7.1 cm (subtending 90°  $\times$  60° visual angle). They were generated on a PC running Ubuntu (version 16.04) using MATLAB and the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997). Participants were instructed to maintain fixation on a white central fixation dot (0.25° diameter) presented on a midgray background of 6.54 cd/m<sup>2</sup>. To aid ocular fusion and maintain stable and vivid depth perception, 16 gray circular placeholders (each 0.8° in diameter) were presented at evenly spaced intervals along an imaginary circle with a radius of 2.5°. The location of the placeholders in depth was either  $-0.1^\circ$  or  $0.1^\circ$ , based on retinal disparity. Depth was varied such that alternating pairs of placeholders had either a positive or a negative disparity (i.e., two close, then two far, then two close, etc.; see Figure 1). Memory-item colors were selected from a circle in CIE La\*b\* color space ( $L = 70$ ,  $a = 20$ ,  $b = 38$ , radius = 60). The two target colors were always 90°  $\pm$  10° apart along the circular color space. We opted to maintain this separation in color space so that the separability of the memory items in color space would remain relatively stable, allowing us to manipulate only 2-D and 3-D spatial separability across experimental conditions. The two memory targets were always presented either close in 2-D space (adjacent, with their centers 0.98° apart) or farther away (centers 2.78° apart), and they could be on the same or different depth planes. This produced four levels of 3-D (same vs. different) and 2-D (close vs. far) separation: same-close, different-close, same-far, and different-far. Note that the two memory targets were always presented in the same hemifield to maximize interitem competition (Alvarez & Cavanagh, 2005; Cohen et al., 2016; Störmer, Alvarez, & Cavanagh, 2014). No color calibration was done on the Oculus goggles. However, since the locations, sizes, and colors of memory items are consistent across all conditions, we believe that any error from calibration will affect all conditions equally. In general, the error introduced by the memory task itself is very large relative to any display properties; reliable data in such paradigms can even be obtained in continuous color-report tasks conducted in entirely uncontrolled settings (e.g., over the internet with all subjects using their own personal computer; Brady & Alvarez, 2015).

On each trial, two colored stimuli were presented for 150 ms and participants had to remember the color of

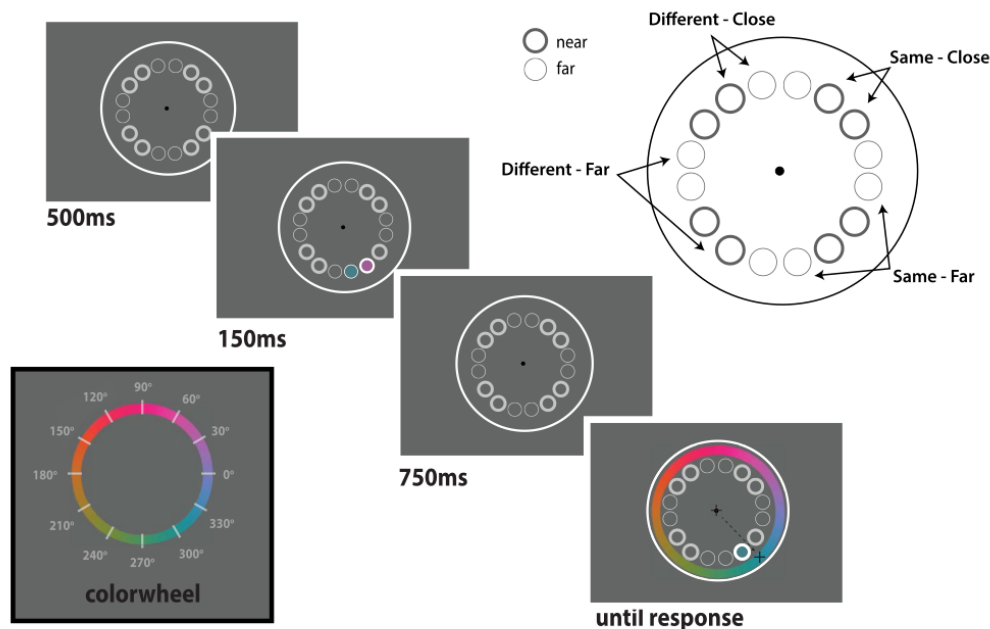


Figure 1. Each trial started with a 500-ms fixation period during which only the 16 placeholders were shown. Here, light and dark circles indicate placeholders on the far and near depth planes, respectively (this is only for visualization purposes—all placeholders were the same shade of gray in the actual experiment). Next, two memory targets were presented for 150 ms, followed by a 750-ms delay. After the delay, a color wheel was presented together with a cue outlining one of the previous target locations, and participants moved the cursor to report the hue previously shown at the cued location. The two target colors were presented in either the same or different depth planes in 3-D coordinates (same vs. different) and either close or far in 2-D space (see insert at top right). The lower left insert shows the color wheel that we used in the experiment.

both during a 750-ms delay period. After the delay, one of the two colors was probed by increasing the thickness of one of the placeholders. Together with the location probe, a color wheel ( $3^\circ$  radius from the center,  $0.5^\circ$  wide, randomly rotated on each trial) and a crosshair appeared over the fixation dot. Participants used the mouse to move the crosshair from the fixation dot to the hue on the color wheel that most closely resembled the color of the probed memory target (Wilken & Ma, 2004). The next trial started after participants clicked the mouse to record their response, and this procedure was repeated 96 times per experimental condition (384 trials in total, conditions randomly interleaved).

### Analyses

We generated a distribution of errors for each participant by computing the difference between the cued target color and the reported color (reported $^\circ$  – target $^\circ$ ) on each trial. To clearly visualize the shape of this error distribution and its relationship to the

nontarget color, we flipped the sign of the error such that the nontarget color was always  $90^\circ$  counterclockwise from the cued target (Figure 2). A commonly used mixture model (Bays et al., 2009; Zhang & Luck, 2008) was fitted to the error distribution under the assumption that responses reflect a mixture of responses to the target color, responses to the nontarget color, and random guesses. This model had four free parameters: the bias ( $b$ , in degrees) of the responses, the standard deviation ( $SD$ ) of the responses (both target and nontarget), the probability of swapping errors ( $s$ , in percent), and the guess rate ( $g$ , in percent; Bays, 2015; Bays et al., 2009; Zhang & Luck, 2008). It was fitted separately to data from each condition for each participant using Mem-Toolbox (Suchow, Brady, Fougne, & Alvarez, 2013). A repeated-measures analysis of variance was then performed to evaluate the impact of 2-D (near/far) and 3-D (same/different depth plane) spatial separation on the estimated model parameters.

It is important to note that the mixture model may have limitations (Schurgin et al., 2018); in particular, precision and guess rate may not be truly separable

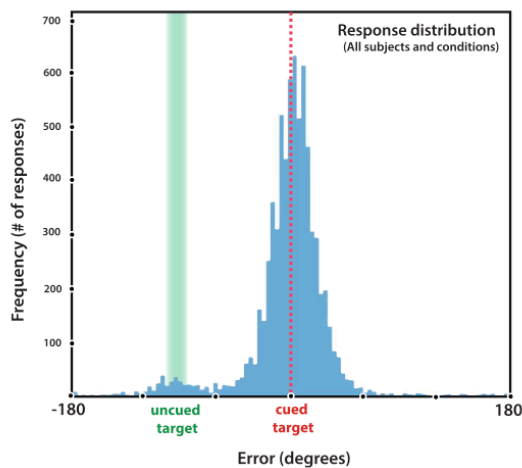


Figure 2. Results of Experiment 1 as a histogram of the responses centered around the target color, shown collapsed across all participants and conditions. The nontarget colors were aligned to approximately  $-90^\circ$  ( $\pm 10^\circ$ ) relative to the target color by flipping the sign of responses on trials where the nontarget was  $+90^\circ$  ( $\pm 10^\circ$ ) relative to the target (note that the width of the shaded green area reflects the  $\pm 10^\circ$  jitter in the uncued target color). Swap errors are apparent from the small bump centered on the nontarget color.

parameters. However, we opted to use the mixture model in this particular experiment because it allowed us to account for systematic biases and for responses to nontargets (swap errors), which are difficult to account for without using a model of the response distribution. For example, without explicit accounting for swap errors, nontarget responses would be treated as  $90^\circ$  errors even though they were actually accurate responses to the nontarget color. However, to check that our results were not dependent on the details of the mixture model, we also performed a post hoc analysis where we developed a nonparametric procedure to quantify memory precision while taking systematic biases and swap errors into account: First, we computed the error (in degrees) of all responses that were centered around the target and nontarget colors (i.e., including responses to nontarget colors as precise responses). Then, in an effort to attenuate the effect of systematic biases, we computed the mean absolute error within  $\pm 60^\circ$  from the peak (mode) of each error-response distribution (i.e., target and nontarget distributions). This allowed us to nonparametrically examine errors without any strong assumptions about the separability of the guess rate and precision parameters of a mixture model.

## Results

Responses were more precise (lower mixture-model *SD*) both when the two memoranda were separated by a greater distance in 2-D spatial position (near/far),  $F(1, 26) = 4.921$ ,  $p = 0.036$ , and when the two memoranda were presented on different depth planes (same/different planes),  $F(1, 26) = 5.677$ ,  $p = 0.025$ , with no interaction between these factors,  $F(1, 26) = 0.06$ ,  $p = 0.808$  (Figure 3A). As shown in Figure 3B, there was a consistent bias such that responses were repelled slightly but consistently away from the nontarget color,  $t(1, 26) = 5.81$ ,  $6.63$ ,  $6.47$ , and  $7.77$ , respectively, for same-close, different-close, same-far, and different-far, all  $p$ s  $< 0.0001$ . However, there was no difference in the magnitude of this bias as a function of separation in 2-D,  $F(1, 26) = 0.002$ ,  $p = 0.965$ , or in 3-D,  $F(1, 26) = 1.377$ ,  $p = 0.251$ , and no interaction between these factors,  $F(1, 26) = 0.983$ ,  $p = 0.331$ . The probability of swapping (i.e., nontarget reports; Figure 3C) did not depend on whether the items were spatially close or far away from each other in 2-D space,  $F(1, 26) = 1.633$ ,  $p = 0.213$ , and there was a nonsignificant trend toward more swap errors when targets were presented on different depth planes,  $F(1, 26) = 3.211$ ,  $p = 0.085$ . No interaction was observed,  $F(1, 26) = 1.889$ ,  $p = 0.181$ . There were also no differences in guess rates estimated by the mixture model across conditions—separation in 2-D:  $F(1, 26) = 0.008$ ,  $p = 0.93$ ; separation in 3-D:  $F(1, 26) = 1.481$ ,  $p = 0.235$ ; interaction:  $F(1, 26) = 0.366$ ,  $p = 0.55$  (Figure 3D).

The quantitative results from this mixture modeling match with the qualitatively observable shapes of the kernel density plots for each condition (Figure 3A–3D vs. 3E, computed using a Gaussian kernel with a standard deviation of  $4^\circ$ ), and the nonparametric analysis of response precision yielded comparable results: The average absolute error around the target was higher when two items were separated both in 2-D,  $F(1, 26) = 6.66$ ,  $p = 0.016$ , and in 3-D,  $F(1, 26) = 6.40$ ,  $p = 0.018$ , and there was no interaction,  $F(1, 26) = 0.46$ ,  $p = 0.505$ .

To evaluate statistical power in our study, we performed a post hoc bootstrapping analysis in which we systematically varied the number of participants. We resampled with replacement data from different numbers of participants, ranging from two to 27, and on each resample we computed the mean differences between conditions. This process was then repeated 1,000 times. On each iteration, we did the same analysis of both the parameters from the mixture model and the nonparametric mean absolute error, and found that both analyses reached stable statistical significance (two-sided  $p < 0.05$ ) with a minimum of 20 participants.

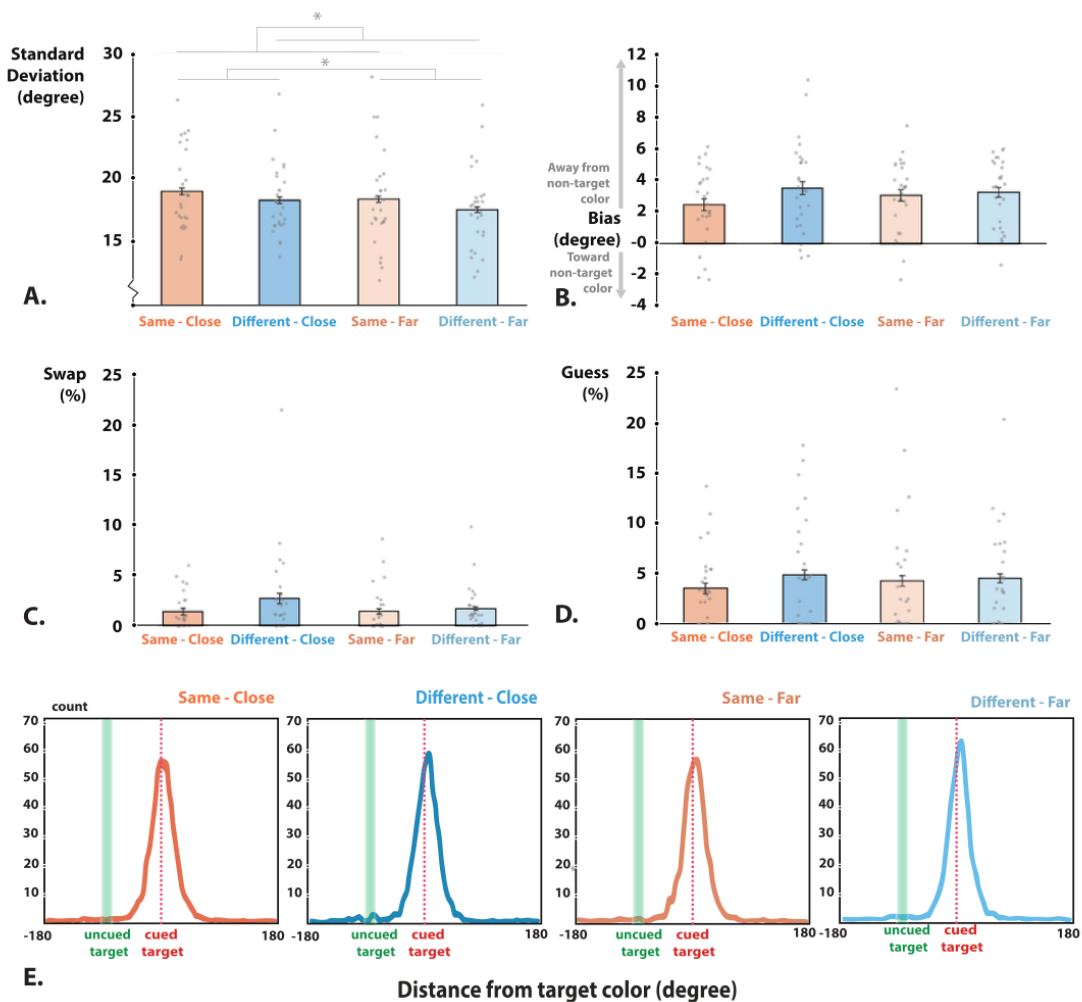


Figure 3. Results of Experiment 1 in terms of the parameters from mixture modeling. (A) The standard deviations are lower when two memory items are spatially far away or when they are on different depth planes (lower standard deviation is associated with higher precision).  $*p < 0.05$ . (B) There are systematic biases away from the nontarget color in all conditions but no significant differences in biases between conditions. (C) There are no significant differences in swap error rate, nor in (D) guess rate. (E) Four kernel density plots of group-level error responses of each condition centered around the target color (from left: same-close, different-close, same-far, and different-far). The shapes of the distributions qualitatively agree with the parameters from the model. Error bars (in A, B, and C,) represent  $\pm 1$  standard error of the mean.

Together these results suggest that spatial separability both within and between different depth planes is associated with higher precision memories in VWM. Importantly, no effects of spatial separability were found on any of the other parameters, suggesting that it is the memory strength that improves once items are separated in either 2-D or 3-D space.

Finally, note that the bias we observed in the target responses was always positive, or away from the nontarget, which is consistent with previous studies showing repulsion biases away from other task-relevant items (Bae & Luck, 2017; Golomb, 2015; Marshak & Sekuler, 1979; Rademaker, Bloem, De Weerd, & Sack, 2015; Rauber & Treue, 1998; Scocchia, Cicchini, &

Triesch, 2013). Interestingly, one study that examined repulsion bias as a function of color similarity between items (Golomb, 2015) showed repulsion biases only when items were close in feature space—specifically less than 60° apart in feature space—while attraction biases were reported when memoranda were more than 60° apart in feature space. However, in the current study we observed repulsion biases even with colors separated by 90° in feature space. Numerous aspects of the current task differed from that previous work (e.g., number of memory items, encoding time, delay time), and many of these factors could affect whether repulsion or attraction is observed in the data and account for the differences between findings.

## Experiment 2

The results from Experiment 1 suggest that separating memoranda within and between depth planes increases memory precision, presumably because interference between the items is reduced. Here we examine the effects of depth on VWM capacity, focusing on the ways it might improve attentional filtering. Studies have shown that the number of items that people can hold in memory with high fidelity may decrease once the number of items to be remembered is large and difficult to manage. For example, one person might be capable of remembering four items with a high degree of fidelity when there are only four to be remembered. However, that same person might remember fewer than four items with a high degree of fidelity when there are 12 memoranda to retain (Cowan & Morey, 2006; Cowan, Morey, AuBuchon, Zwilling, & Gilchrist, 2010; Cusack, Lehmann, Veldsman, & Mitchell, 2009; Linke, Vicente-Grabovetsky, Mitchell, & Cusack, 2011; Vogel, McCollough, & Machizawa, 2005). This phenomenon has usually been attributed to a failure of attentional filtering, as trying to store everything in the display may have negative consequences. Previous work has shown that spatial location can aid attentional filtering (Vogel et al., 2005). Therefore, we hypothesized that separating items in depth might also aid attentional filtering. In particular, we predicted that once participants have a large number of items to remember and therefore must rely on attentional filtering to select a subset of items to represent with high fidelity, separation in depth should promote a higher memory capacity. Alternatively, it is possible that increasing the number of memory items in a 3-D display might lead to poorer overall performance due to an increased demand to distribute spatial attention across a larger volume of space. To test these accounts, we manipulated memory set size across a range from two to 12 items. We also independently

assessed each participant's ability to exploit stereo depth cues so that we could evaluate the relationship between the salience of depth information and its impact on VWM capacity across participants.

## Methods

### Participants

A new set of 22 healthy volunteers (14 women, eight men; mean age = 19.67 ± 0.45 years) from the UCSD community participated in the experiment. All procedures were approved by the UCSD Institutional Research Board. All participants reported normal or corrected-to-normal vision without color-vision deficiency and provided written informed consent. Participants were unaware of the purpose of the study and received course credits or monetary compensation for their time (\$10/hr). All participants passed the same stereovision test used in Experiment 1, and none were excluded.

### Stimuli and procedure

Unless otherwise mentioned, stimulus generation and presentation were identical to Experiment 1. The main VWM task in Experiment 2 (Figure 4A) used a delayed-match-to-sample paradigm. At the beginning of each trial, 12 placeholders were presented (each 1° in diameter, presented at 2.5° from fixation) for 500 ms. The depth separation of the placeholders was experimentally manipulated: On 50% of trials, they were all presented on the same depth plane (all on the near plane on 25%, all on the far plane on another 25%)—the same-depth condition. On the remaining 50% of trials, half of the placeholders were on the near plane and the other half were on the far plane—the different-depths condition. Next, two, four, six, eight, or 12 colored memory targets were briefly presented (500 ms) at a random subset of the 12 placeholders, with the restriction that in the different-depths condition half of the items were assigned to near and the other half to far placeholders (12 stimuli were shown in every placeholder). Colors were randomly chosen from a set of 12 unique colors. After a 900-ms delay, a single test color was presented at one of the memory-target locations, either matching or not matching the target color previously shown at that location. Participants indicated match or nonmatch by pressing the X or C key, respectively, with matches occurring on 50% of trials and nonmatches created by placing one of the other remembered items from the initial display in the test location. For each participant, we collected 80 trials for each set-size (2, 4, 6, 8, and 12) and depth condition (same vs. different depths), leading to 800 total trials. Participants performed 10 blocks of 80 trials each, with

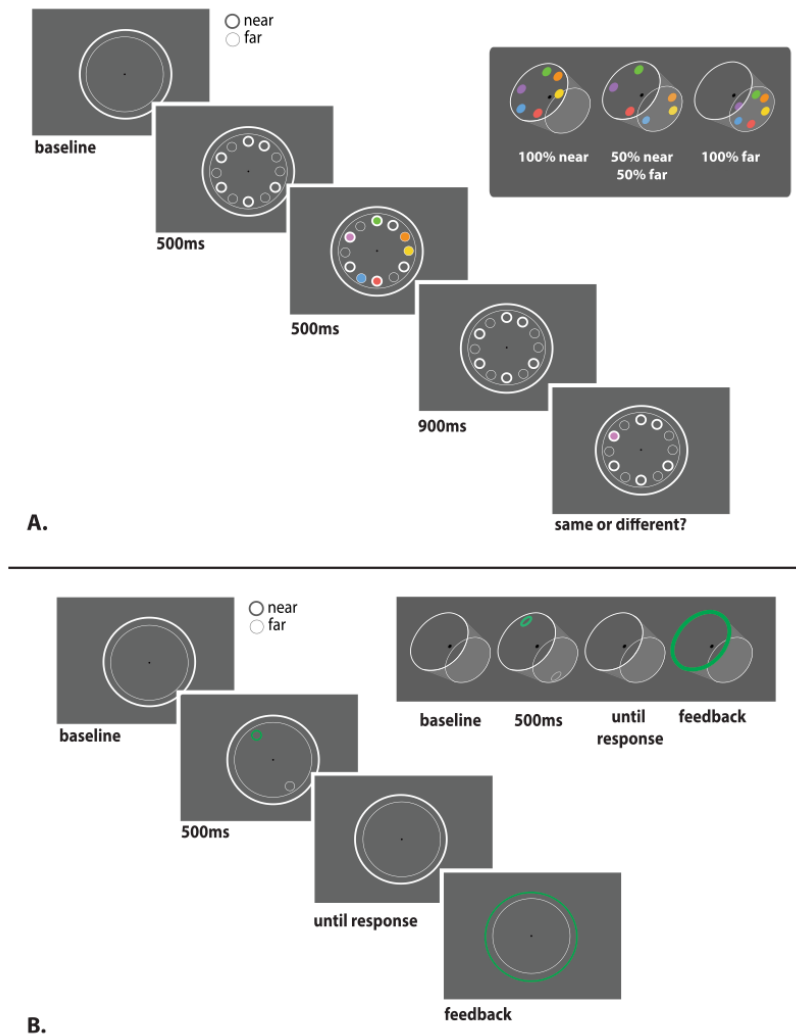


Figure 4. Experimental procedure for Experiment 2. (A) In this single-probe change-detection paradigm, each trial started with the presentation of 12 placeholders. Placeholders could have one of three possible depth relationships: all on the near depth plane, all on the far depth plane, or half on the near and the other half on the far depth plane. After 500 ms, two, four, six, eight, or 12 colored memory items were presented for 500 ms, followed by a 900-ms delay period. Next, a single test item was presented at a location previously occupied by one of the memory items, and participants indicated whether the color of the test was the same as or different from the color of the memory target previously shown at that location. (B) The independent depth-discrimination task. On each trial, two placeholders briefly appeared, each on a different depth plane. Participants indicated whether the target (in green) was on the near or far plane. Performance on this task was used as an indicator of how well participants could perceive depth using our stereo-display setup.

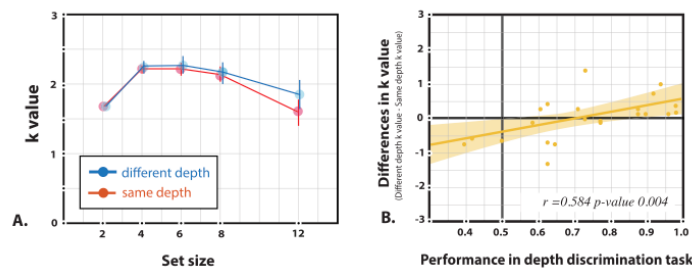


Figure 5. Main results of Experiment 2. (A) Visual-working-memory capacity (Cowan's  $k$ ) as a function of set size. There were no differences in capacity when memory items were displayed on planes at the same (red) or different (blue) depths. Observed changes in  $k$  as a function of set size are consistent with previous studies (Cowan & Morey, 2006). (B) The impact of depth separation (on the y-axis) was calculated by taking the capacity  $k$  for items presented on different depth planes minus the  $k$  for items presented on the same depth plane. Thus, larger numbers indicate a larger benefit of presenting items separated in depth. The ability of participants to discriminate the two depth planes in our experimental setup (on the x-axis) was positively correlated with the benefits they gained from items presented on different depth planes. Shaded regions indicate  $\pm 1$  standard error of the mean.

each block lasting  $\sim 5$  min. Note that using a delayed-match-to-sample paradigm required less time per trial than continuous report and thus allowed us to quickly evaluate memory performance across five set sizes for items on the same and different depth planes.

To enable us to evaluate how well participants could perceive memoranda presented on the two different depth planes, participants also completed a 48-trial depth-discrimination task (Figure 4B) prior to participating in the main task. During this independent task, two placeholders were presented for 500 ms, with one on the near plane and the other on the far plane (with respect to fixation). The location of the two placeholders was chosen at random from the 12 possible locations used in the main task. Participants had to indicate whether a target (specified by a green circle outline) was on the near or far plane. The ability of each participant to accurately identify the correct depth plane in this task was used to predict the benefits of the depth information during the VWM task.

### Analyses

We estimated each participant's VWM capacity using a standard measure appropriate for single-probe change detection—Cowan's  $k$  (Cowan, 2010; Pashler, 1988)—as follows:

$$k = (\text{hit rate} - \text{false alarm}) \times \text{set size}.$$

As in Experiment 1, repeated-measures analyses of variance were used for the main analyses. Additionally, the impact of a participant's ability to perceive the stimuli in depth (measured with the independent depth-discrimination task) on performance during the VWM task was assessed using correlational analyses.

### Results

There was a significant main effect of set size on observed  $k$  values,  $F(4, 84) = 5.26$ ,  $p < 0.001$  (Figure 5A), such that estimates of capacity were lower for very small and very large set sizes; a linear fit failed to capture a significant amount of variance,  $F(1, 215) = 0.59$ ,  $p = 0.44$ , while adding a quadratic significantly improved the fit,  $F(3, 215) = 3.81$ ,  $p = 0.011$ . However, there was no effect of depth condition,  $F(1, 21) = 0.018$ ,  $p = 0.895$ , and no Set size  $\times$  Depth condition interaction,  $F(4, 84) = 0.107$ ,  $p = 0.98$ . While this may suggest that presenting memory items on the same versus different depth planes did not affect memory capacity, we found a positive correlation between depth-discrimination ability (as indexed during the independent depth-discrimination task) and the impact of separation in depth (as manipulated in the main VWM task). Specifically, participants with better stereo depth perception showed a larger performance benefit when items were presented on different depth planes (Pearson's  $r = 0.58$ ,  $p = 0.004$ ; Figure 5B), and this correlation was still significant when participants with negative  $k$  values were excluded from the analysis (Pearson's  $r = 0.55$ ,  $p = 0.012$ ). This effect was systematically related to set size, such that correlations grew stronger as set size increased (Figure 6, bottom row)—set size 2:  $r < 0.0001$ ,  $p = 0.99$ ; set size 4:  $r = -0.05$ ,  $p = 0.81$ ; set size 6:  $r = 0.38$ ,  $p = 0.08$ ; set size 8:  $r = 0.42$ ,  $p = 0.05$ ; set size 12:  $r = 0.54$ ,  $p = 0.008$ .

Importantly, the correlations between the depth-discrimination task and VWM performance were found selectively in the 3-D condition (Pearson's  $r = 0.49$ ,  $p = 0.05$ ) but were not found in the 2-D condition (Pearson's  $r = 0.05$ ,  $p = 0.80$ ). The correlation analyses after excluding two subjects with negative average  $k$

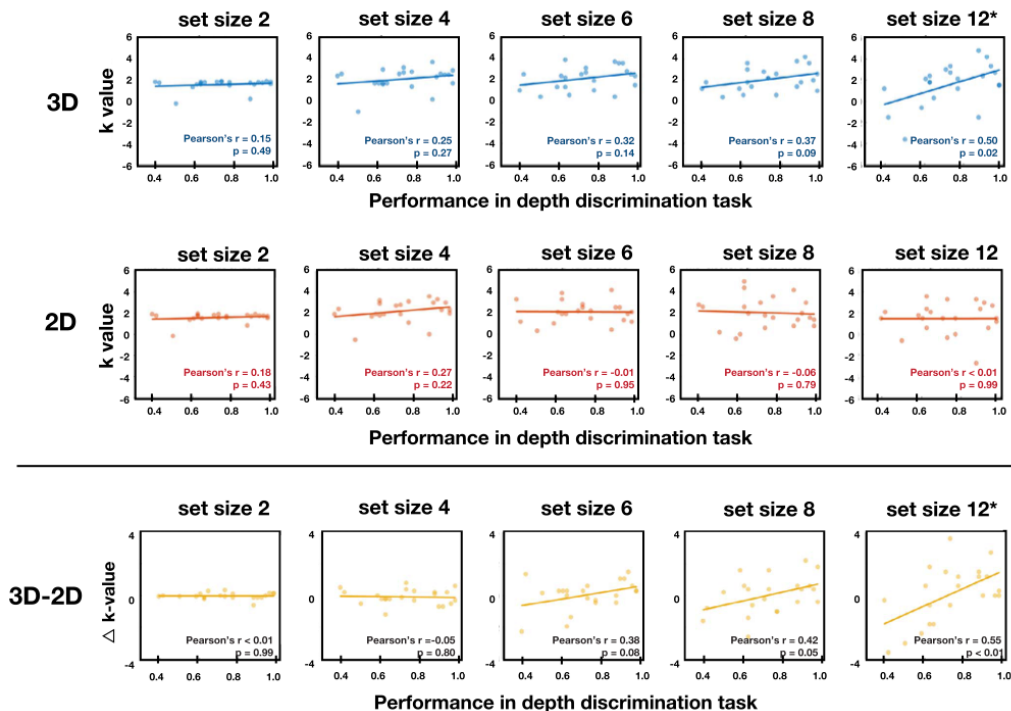


Figure 6. The degree of positive correlation between depth-discrimination ability (on the x-axis) and performance on the visual-working-memory task (on the y-axis). Participants who performed better on the depth-discrimination task also performed better on the visual-working-memory task at larger set sizes, but only when the memoranda were on different depth planes (upper row). There was no correlation between performance on the depth-discrimination task and on the visual-working-memory task when the memoranda were in the same depth plane (middle row). The benefit associated with having the memoranda separated into different depth planes (difference in  $k$  value on the y-axis) grew stronger as set size increased (bottom row in panels).

values found similar results (3-D: Pearson's  $r = 0.49$ ,  $p = 0.028$ ; 2-D: Pearson's  $r = -0.008$ ,  $p = 0.97$ ). We ran a dependent correlation test and found a significant difference between the 2-D and 3-D correlations,  $t = 3.08$ ,  $p = 0.01$ , showing that the 3-D correlations were reliably higher than in the 2-D condition. This indicates that the correlation was not related to differences in general arousal or motivation (Figure 6). We believe that the effect is robust given that these correlations grow monotonically stronger as set size increases. To ensure that this analysis had enough power, we did a bootstrapping analysis in which we resampled data from a different number of participants (between five and 22) with replacement 1,000 times (just as we did in Experiment 1). We found stable positive correlations (more than 97.5% of the simulations had positive correlations; equal to two-sided  $p < 0.05$ ) when there were at least 10 participants included.

As an alternate means of assessing the data, we sorted participants into two groups based on a median split of their depth-discrimination ability as assessed using the independent task (Figure 7). We found a main effect of set size,  $F(4, 80) = 5.22$ ,  $p < 0.001$ , but not of depth plane,  $F(1, 20) = 0.03$ ,  $p = 0.87$ . There was also a significant two-way interaction such that separation in depth led to improved performance only for those subjects who performed well on the independent depth-discrimination task,  $F(1, 20) = 10.95$ ,  $p = 0.004$ . Performance on the depth-discrimination task was not associated with an overall change in VWM performance levels collapsed across set size and condition,  $F(1, 20) = 0.79$ ,  $p = 0.39$ , suggesting that the two groups of subjects were equally motivated to perform the task. Nevertheless, there was a three-way interaction such that participants who performed well on the independent depth task showed the benefit of depth at larger set size,  $F(4, 80) = 3.622$ ,  $p = 0.009$ .



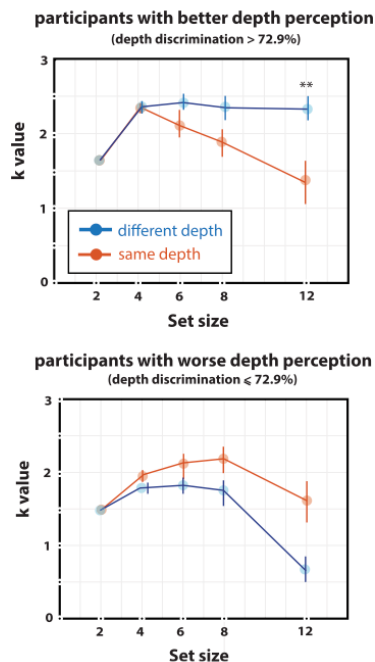


Figure 7. Participants who exhibited better depth discrimination (upper panel), based on a median split of performance in the independent depth-discrimination task, benefited more from the presence of depth information, particularly at high set sizes.  $**p < 0.01$ . The error bars represent  $\pm 1$  standard error of the mean. For participants who exhibited worse depth discrimination (lower graph), the  $k$  value appeared to be lower when memoranda were on different depth planes, but this did not reach significance. Note that the performance from both groups was comparable when the memoranda were on the same depth plane (compare red lines between the two panels).

To follow up on these findings, we also performed post hoc tests separately on data within the low and high depth discriminators. We found that the high depth discriminators did better on the VWM task when the items were separated in depth—main effect:  $F(1, 11) = 6.79, p = 0.024$ —especially with larger set sizes—interaction:  $F(4, 44) = 3.53, p = 0.014$ . This indicates that participants with better depth perception ( $> 72.9\%$  accuracy) performed better on different-depths displays, but only at larger set sizes (Figure 7, top panel)—set size 2:  $t(1, 11) = -0.25, p = 0.81$ ; set size 4:  $t = 0.06, p = 0.96$ ; set size 6:  $t = 1.83, p = 0.09$ ; set size 8:  $t = 1.44, p = 0.18$ ; set size 12:  $t = 2.78, p = 0.02$ . For the low depth discriminators there was a small opposite trend such that performance was lower when memoranda were in different depth planes. However, the analysis of

variance did not reveal a significant main effect of separation in depth,  $F(1, 9) = 4.439, p = 0.064$ , nor an interaction,  $F(4, 36) = 1.052, p = 0.394$ . And post hoc paired  $t$  tests were also nonsignificant (Figure 7, bottom panel)—set size 2:  $t(1, 9) = -0.35, p = 0.73$ ; set size 4:  $t = -1.35, p = 0.21$ ; set size 6:  $t = -0.78, p = 0.46$ ; set size 8:  $t = -1.14, p = 0.29$ ; set size 12:  $t = -1.63, p = 0.14$ .

We also performed post hoc tests separately on data from same- and different-depth conditions. Importantly, there was an interaction between low and high depth discriminators and set size when the memoranda were on different planes,  $F(4, 80) = 2.87, p = 0.028$ , but not when they were on the same plane,  $F(4, 80) = 0.75, p = 0.564$ , indicating that the benefits of better depth perception were restricted to trials where the memory load was high and memoranda were presented in separate depth planes. Moreover, the lack of an effect of depth-perception ability on performance in the same-depth condition further suggests that differences in overall motivation between the two groups of participants cannot account for the observed differences in the different-depths condition.

## Discussion

Perceiving the world in 3-D is a seemingly effortless endeavor, and depth information is fundamental to perceptual organization of the visual world into objects and surfaces, as well as guiding motor interactions with objects in the environment. However, the manner in which the visual system represents in-plane 2-D information versus 3-D depth information is fundamentally different. First, depth information must be indirectly inferred based on operations applied to the 2-D input provided by the projection of light onto the retina. Thus, depth is a second-order feature of visual representation that is indirectly constructed from a set of binocular and monocular cues. Second, the visual system is organized such that ordinal information about the 2-D layout of a visual scene is preserved: Stimuli that are closer to each other in the world are represented by neurons that are closer to each other in the retina and in later visual areas. In contrast, the extent of topographic representations of depth in visual cortex is not well understood, with only a few recent studies suggesting that a structured layout of depth exists in some visual areas (Finlayson et al., 2017). Here we show that separating memoranda in both the 2-D plane and 3-D depth improves VWM performance, consistent with the idea that separating stimuli in depth attenuates interitem competition and interference which affects how people perceive the display (Ander- sen, 1990; Finlayson & Golomb, 2016; Kooi, Toet, Tripathy, & Levi, 1994; Lehmkuhle & Fox, 1980;

Papathomas, Feher, Julesz, & Zeevi, 1996). This is also in line with evidence that people remember real-world 3-D objects better than drawings or photographs of the same objects, even when retinal images are roughly matched (Snow, Skiba, Coleman, & Berryhill, 2014). Furthermore, separating memoranda in depth had the biggest impact on performance when set size increased, suggesting that at least some participants were able to exploit this additional 3-D spatial information to help encode and maintain distinct representations of remembered items.

Previous work has produced mixed results regarding the impact of depth on VWM. For example, two recent studies using a change-detection task did not find any effect of separating memoranda in depth using a display in which all items were presented simultaneously (Qian et al., 2017; Reeves & Lei, 2014). An earlier study also found no benefits of depth using a simultaneous display, but did find that participants had a higher VWM capacity under stereoscopic viewing conditions when each item was presented sequentially on a different depth plane (Xu & Nakayama, 2007). The authors of this latter study hypothesized that perceiving items separated in depth might be inherently more difficult in a simultaneous display, as participants need to attend more than one depth plane at a time—in sequential displays this is presumably no longer an issue, unveiling the benefits of separation in depth. Interestingly, that same study showed that separation in depth had a benefit above and beyond other grouping cues, like changing the configuration of the memoranda by grouping subsets of memoranda into squares or circles. However, in everyday life we perceive depth information in stable and whole scenes, not in sequence. Because sequential presentation of depth information is one step removed from real-world conditions, it thus remains unclear from this previous work whether separation in depth yields any benefit without separation in time.

One alternative explanation for previous results which did not find a benefit to depth when using simultaneous displays is that participants simply differ in terms of how well they perceive the depth cues used in the experimental displays. In our Experiment 2, we independently measured individual differences in depth perception and found a clear benefit for separating memoranda in depth within the group of participants who were better able to exploit stereo cues to support depth perception. It is important to note that our depth-discrimination task required participants to be able to rapidly acquire depth information in order to accurately parse the array. Thus, even though all of the participants passed a basic stereovision screening test, there were still large individual differences in how efficiently they perceived depth information at the relatively brief exposure duration (i.e., 500 ms) used in

the depth-perception and VWM tasks. For example, participants who have stereovision but did poorly on the depth-perception task might not be able to rapidly switch their attention between depth planes (or not be able to simultaneously attend to both depth planes), resulting in relatively worse performance in the 3-D condition of the VWM task. The results from Experiment 2 also showed greater benefits of separation in depth at larger set sizes, consistent with the idea that separation in depth attenuates interitem competition and possibly improves attentional filtering. As visual attention (the ability to selectively process visual information) and VWM (the ability to retain visual information) are related cognitive mechanisms, one possibility is that the separation of items in depth affects how visual attention is distributed (e.g., sequential focal attention rather than simultaneous more distributed attention). Consequently, interference (and thus error) could be reduced, the difference between items amplified (two colors seen or remembered as more different; e.g., Finlayson & Golomb, 2016), and the relative position of items partially lost (more swap errors, e.g., mean nontarget responses of 19% vs. 4% in sequential vs. simultaneous display; Gorgoraptis, Catalao, Bays, & Husain, 2011).

It remains an open question to what extent our results arise from differences in binocular disparity per se, differences in perceived depth, or more general properties of surface perception (e.g., Nakayama, He, & Shimojo, 1995) regardless of the cues that give rise to such surfaces. Some work has suggested that perceptual benefits in related tasks are a result of binocular disparity rather than depth (Finlayson & Golomb, 2016), whereas many recognition tasks seem to largely benefit from coherent surface organization rather than binocular disparity (Nakayama, Shimojo, & Silverman, 1989). Future research will be needed to dissociate these different factors and their respective influences on VWM performance.

In summary, the present results demonstrate that separating memoranda in depth improves visual working memory. In Experiment 1, we show that separation in depth benefits VWM on a scale similar to separating memoranda in 2-D. The similarity of these depth effects to effects observed with 2-D space is particularly interesting given that spatial and depth information are fundamentally different, with 2-D information encoded directly at the retina while 3-D information needs to be indirectly inferred based on binocular and monocular cues. In Experiment 2, we show further that separation in depth confers the largest benefits when participants are better at exploiting stereo depth cues and when interitem competition is highest due to larger set sizes. Together, these observations suggest that interitem interference can occur after the computation of second-order properties

of the visual scene and not just at the level of retinotopically organized representations reflecting 2-D in-plane separation. Showing items at varying depths may thus confer an important benefit to behavioral performance in psychophysical tasks.

*Keywords:* visual working memory, memory biases, depth perception

## Acknowledgments

This work was supported by NEI R01-EY025872 and a James S. McDonnell Foundation Scholar Award to JTS, a Thai Red Cross Society grant to CC, by the European Union's Horizon 2020 research and innovation program under Marie Skłodowska-Curie Grant Agreement No. 743941 to RLR, and by NSF CAREER Award No. BCS-1653457 to TFB.

Commercial relationships: none.

Corresponding author: Chaipat Chunharas.

Email: chaipat.c@chula.ac.th.

Address: Department of Medicine, King Chulalongkorn Memorial hospital, Bangkok, Thailand.

## References

- Alvarez, G. A., & Cavanagh, P. (2005). Independent resources for attentional tracking in the left and right visual hemifields. *Psychological Science*, *16*(8), 637–643.
- Andersen, G. J. (1990). Focused attention in three-dimensional space. *Perception & Psychophysics*, *47*(2), 112–120.
- Andersen, G. J., & Kramer, A. F. (1993). Limits of focused attention in three-dimensional space. *Perception & Psychophysics*, *53*(6), 658–667.
- Atchley, P., Kramer, A. F., Andersen, G. J., & Theeuwes, J. (1997). Spatial cuing in a stereoscopic display: Evidence for a “depth-aware” attentional focus. *Psychonomic Bulletin & Review*, *4*(4), 524–529.
- Baddeley, A. (1986). Working memory, reading and dyslexia. In E. Hjelmquist & L.-G. Nilsson (Eds.), *Advances in Psychology* (Vol. 34, pp. 141–152). Amsterdam, the Netherlands: North-Holland.
- Bae, G.-Y., & Luck, S. J. (2017). Interactions between visual working memory representations. *Attention, Perception & Psychophysics*, *79*(8), 2376–2395, <https://doi.org/10.3758/s13414-017-1404-8>.
- Bays, P. (2015). Evaluating and excluding swap errors in analogue report. *Journal of Vision*, *15*(12): 675, <https://doi.org/10.1167/15.12.675>. [Abstract]
- Bays, P. M. (2016). Evaluating and excluding swap errors in analogue tests of working memory. *Scientific Reports*, *6*, 19203.
- Bays, P. M., Catalao, R. F. G., & Husain, M. (2009). The precision of visual working memory is set by allocation of a shared resource. *Journal of Vision*, *9*(10):7, 1–11, <https://doi.org/10.1167/9.10.7>. [PubMed] [Article]
- Bays, P. M., Gorgoraptis, N., Wee, N., Marshall, L., & Husain, M. (2011). Temporal dynamics of encoding, storage, and reallocation of visual working memory. *Journal of Vision*, *11*(10):6, 1–15, <https://doi.org/10.1167/11.10.6>. [PubMed] [Article]
- Bays, P. M., & Husain, M. (2008, August 8). Dynamic shifts of limited working memory resources in human vision. *Science*, *321*(5890), 851–854.
- Bays, P. M., Wu, E. Y., & Husain, M. (2011). Storage and binding of object features in visual working memory. *Neuropsychologia*, *49*(6), 1622–1631.
- Brady, T. F., & Alvarez, G. A. (2011). Hierarchical encoding in visual working memory: Ensemble statistics bias memory for individual items. *Psychological Science*, *22*(3), 384–392.
- Brady, T. F., & Alvarez, G. A. (2015). Contextual effects in visual working memory reveal hierarchically structured memory representations. *Journal of Vision*, *15*(15):6, 1–24, <https://doi.org/10.1167/15.15.6>. [PubMed] [Article]
- Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4), 433–436.
- Cohen, M. A., Rhee, J. Y., & Alvarez, G. A. (2016). Limits on perceptual encoding can be predicted from known receptive field properties of human visual cortex. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(1), 67–77.
- Conway, A. R. A., Cowan, N., Bunting, M. F., Theriault, D. J., & Minkoff, S. R. B. (2002). A latent variable analysis of working memory capacity, short-term memory capacity, processing speed, and general fluid intelligence. *Intelligence*, *30*(2), 163–183.
- Cowan, N. (2010). The magical mystery four: How is working memory capacity limited, and why? *Current Directions in Psychological Science*, *19*(1), 51–57.
- Cowan, N., & Morey, C. C. (2006). Visual working memory depends on attentional filtering. *Trends in Cognitive Sciences*, *10*(4), 139–141.
- Cowan, N., Morey, C. C., AuBuchon, A. M., Zwilling,

- C. E., & Gilchrist, A. L. (2010). Seven-year-olds allocate attention like adults unless working memory is overloaded. *Developmental Science*, *13*(1), 120–133.
- Cusack, R., Lehmann, M., Veldsman, M., & Mitchell, D. J. (2009). Encoding strategy and not visual working memory capacity correlates with intelligence. *Psychonomic Bulletin & Review*, *16*(4), 641–647.
- D'Esposito, M., & Postle, B. R. (2015). The cognitive neuroscience of working memory. *Annual Review of Psychology*, *66*, 115–142.
- Downing, C., & Pinker, S. (1985). *Attention and performance*. Mahwah, NJ: Earlbaum.
- Emrich, S. M., & Ferber, S. (2012). Competition increases binding errors in visual working memory. *Journal of Vision*, *12*(4):12, 1–16, <https://doi.org/10.1167/12.4.12>. [PubMed] [Article]
- Emrich, S. M., Riggall, A. C., Larocque, J. J., & Postle, B. R. (2013). Distributed patterns of activity in sensory cortex reflect the precision of multiple items maintained in visual short-term memory. *The Journal of Neuroscience*, *33*(15), 6516–6523.
- Engel, S. A., Glover, G. H., & Wandell, B. A. (1997). Retinotopic organization in human visual cortex and the spatial precision of functional MRI. *Cerebral Cortex*, *7*(2), 181–192.
- Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. (1999). Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology: General*, *128*(3), 309–331.
- Enns, J. T., & Rensink, R. A. (1990). Sensitivity to three-dimensional orientation in visual search. *Psychological Science*, *1*(5), 323–326.
- Finlayson, N. J., & Golomb, J. D. (2016). Feature-location binding in 3D: Feature judgments are biased by 2D location but not position-in-depth. *Vision Research*, *127*, 49–56.
- Finlayson, N. J., & Grove, P. M. (2015). Visual search is influenced by 3D spatial layout. *Attention, Perception & Psychophysics*, *77*(7), 2322–2330.
- Finlayson, N. J., Remington, R. W., Retell, J. D., & Grove, P. M. (2013). Segmentation by depth does not always facilitate visual search. *Journal of Vision*, *13*(8):11, 1–14, <https://doi.org/10.1167/13.8.11>. [PubMed] [Article]
- Finlayson, N. J., Zhang, X., & Golomb, J. D. (2017). Differential patterns of 2D location versus depth decoding along the visual hierarchy. *NeuroImage*, *147*, 507–516.
- Fukuda, K., Vogel, E., Mayr, U., & Awh, E. (2010). Quantity, not quality: The relationship between fluid intelligence and working memory capacity. *Psychonomic Bulletin & Review*, *17*(5), 673–679.
- Golomb, J. D. (2015). Divided spatial attention and feature-mixing errors. *Attention, Perception & Psychophysics*, *77*(8), 2562–2569.
- Gorgoraptis, N., Catalao, R. F. G., Bays, P. M., & Husain, M. (2011). Dynamic updating of working memory resources for visual objects. *The Journal of Neuroscience*, *31*(23), 8502–8511.
- Grill-Spector, K., & Malach, R. (2004). The human visual cortex. *Annual Review of Neuroscience*, *27*, 649–677.
- Harrison, S. A., & Tong, F. (2009, April 2). Decoding reveals the contents of visual working memory in early visual areas. *Nature*, *458*(7238), 632–635.
- He, Z. J., & Nakayama, K. (1995). Visual attention to surfaces in three-dimensional space. *Proceedings of the National Academy of Sciences, USA*, *92*(24), 11155–11159.
- Hollingworth, A. (2007). Object-position binding in visual memory for natural scenes and object arrays. *Journal of Experimental Psychology: Human Perception and Performance*, *33*(1), 31–47.
- Hollingworth, A., & Rasmussen, I. P. (2010). Binding objects to locations: The relationship between object files and visual working memory. *Journal of Experimental Psychology: Human Perception and Performance*, *36*(3), 543–564.
- Jiang, Y., Olson, I. R., & Chun, M. M. (2000). Organization of visual short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(3), 683–702.
- Kooi, F. L., Toet, A., Tripathy, S. P., & Levi, D. M. (1994). The effect of similarity and duration on spatial interaction in peripheral vision. *Spatial Vision*, *8*(2), 255–279.
- Lehmkühle, S., & Fox, R. (1980). Effect of depth separation on metacontrast masking. *Journal of Experimental Psychology: Human Perception and Performance*, *6*(4), 605–621.
- Linke, A. C., Vicente-Grabovetsky, A., Mitchell, D. J., & Cusack, R. (2011). Encoding strategy accounts for individual differences in change detection measures of VSTM. *Neuropsychologia*, *49*(6), 1476–1486.
- Luck, S. J., & Vogel, E. K. (1997, November 20). The capacity of visual working memory for features and conjunctions. *Nature*, *390*(6657), 279–281.
- Ma, W. J., Husain, M., & Bays, P. M. (2014). Changing concepts of working memory. *Nature Neuroscience*, *17*(3), 347–356.

- Marshak, W., & Sekuler, R. (1979, September 28). Mutual repulsion between moving visual targets. *Science*, 205(4413), 1399–1401.
- Maunsell, J. H., & Newsome, W. T. (1987). Visual processing in monkey extrastriate cortex. *Annual Review of Neuroscience*, 10, 363–401.
- McCarley, J. S., & He, Z. J. (2001). Sequential priming of 3-D perceptual organization. *Perception & Psychophysics*, 63(2), 195–208.
- Nakayama, K., He, Z. J., & Shimojo, S. (1995). Visual surface representation: A critical link between lower-level and higher-level vision. *Visual cognition: An invitation to cognitive science* (Vol. 2, pp. 1–70). Cambridge, MA: MIT press.
- Nakayama, K., Shimojo, S., & Silverman, G. H. (1989). Stereoscopic depth: Its relation to image segmentation, grouping, and the recognition of occluded objects. *Perception*, 18(1), 55–68.
- Nakayama, K., & Silverman, G. H. (1986, March 20). Serial and parallel processing of visual feature conjunctions. *Nature*, 320(6059), 264–265.
- Olson, I. R., & Marshuetz, C. (2005). Remembering “what” brings along “where” in visual working memory. *Perception & Psychophysics*, 67(2), 185–194.
- Panichello, M. F., DePasquale, B., Pillow, J. W., & Buschman, T. (2018). Error-correcting dynamics in visual working memory. *bioRxiv*. Retrieved from <https://www.biorxiv.org/content/early/2018/05/10/319103.abstract>
- Papathomas, T. V., Feher, A., Julesz, B., & Zeevi, Y. (1996). Interactions of monocular and cyclopean components and the role of depth in the Ebbinghaus illusion. *Perception*, 25(7), 783–795.
- Pashler, H. (1988). Familiarity and visual change detection. *Perception & Psychophysics*, 44(4), 369–378.
- Pasternak, T., & Greenlee, M. W. (2005). Working memory in primate sensory systems. *Nature Reviews Neuroscience*, 6(2), 97–107.
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, 10(4), 437–442.
- Pertsov, Y., & Husain, M. (2014). The privileged role of location in visual working memory. *Attention, Perception & Psychophysics*, 76(7), 1914–1924.
- Phillips, W. A. (1974). On the distinction between sensory storage and short-term visual memory. *Perception & Psychophysics*, 16(2), 283–290.
- Postle, B. R., Awh, E., Serences, J. T., Sutterer, D. W., & D’Esposito, M. (2013). The positional-specificity effect reveals a passive-trace contribution to visual short-term memory. *PLoS One*, 8(12), e83483.
- Qian, J., Li, J., Wang, K., Liu, S., & Lei, Q. (2017). Evidence for the effect of depth on visual working memory. *Scientific Reports*, 7(1), 6408.
- Rademaker, R. L., Bloem, I. M., De Weerd, P., & Sack, A. T. (2015). The impact of interference on short-term memory for visual orientation. *Journal of Experimental Psychology: Human Perception and Performance*, 41(6), 1650–1665.
- Rademaker, R.L., Chunharas, C., & Serences, J. T. (2018). Simultaneous representation of sensory and mnemonic information in human visual cortex. *bioRxiv*, <https://doi.org/10.1101/339200>.
- Rademaker, R. L., Park, Y. E., & Sack, A. T. (2018). Evidence of gradual loss of precision for simple features and complex objects in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance*, 44(6), 925–940. Retrieved from <http://psycnet.apa.org/record/2018-08188-001>
- Rademaker, R. L., Tredway, C., & Tong, F. (2012). Introspective judgments predict the precision and likelihood of successful maintenance of visual working memory. *Journal of Vision*, 12(13):21, 1–13, <https://doi.org/10.1167/12.13.21>. [PubMed] [Article]
- Rauber, H. J., & Treue, S. (1998). Reference repulsion when judging the direction of visual motion. *Perception*, 27(4), 393–402.
- Reeves, A., & Lei, Q. (2014). Is visual short-term memory depthful? *Vision Research*, 96, 106–112.
- Schurgin, M. W., Wixted, J. T., & Brady, T. F. (2018). Psychological scaling reveals a single parameter framework for visual working memory. *bioRxiv*. Retrieved from <https://www.biorxiv.org/content/biorxiv/early/2018/05/18/325472.full.pdf>
- Scocchia, L., Cicchini, G. M., & Triesch, J. (2013). What’s “up”? Working memory contents can bias orientation processing. *Vision Research*, 78, 46–55.
- Serences, J. T. (2016). Neural mechanisms of information storage in visual short-term memory. *Vision Research*, 128, 53–67.
- Serences, J. T., Ester, E. F., Vogel, E. K., & Awh, E. (2009). Stimulus-specific delay activity in human primary visual cortex. *Psychological Science*, 20(2), 207–214.
- Sereno, M. I., Dale, A. M., Reppas, J. B., Kwong, K. K., Belliveau, J. W., Brady, T. J., . . . Tootell, R. B. (1995, May 12). Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. *Science*, 268(5212), 889–893.

- Sereno, M. I., Pitzalis, S., & Martinez, A. (2001, November 9). Mapping of contralateral space in retinotopic coordinates by a parietal cortical area in humans. *Science*, *294*(5545), 1350–1354.
- Shin, H., Zou, Q., & Ma, W. J. (2017). The effects of delay duration on visual working memory for orientation. *Journal of Vision*, *17*(14):10, 1–24, <https://doi.org/10.1167/17.14.10>. [PubMed] [Article]
- Simons, D. J., & Levin, D. T. (1997). Change blindness. *Trends in Cognitive Sciences*, *1*(7), 261–267.
- Snow, J. C., Skiba, R. M., Coleman, T. L., & Berryhill, M. E. (2014). Real-world objects are more memorable than photographs of objects. *Frontiers in Human Neuroscience*, *8*, 837.
- Sprague, T. C., Ester, E. F., & Serences, J. T. (2014). Reconstructions of information in visual spatial working memory degrade with memory load. *Current Biology*, *24*(18), 2174–2180.
- Sreenivasan, K. K., Curtis, C. E., & D’Esposito, M. (2014). Revisiting the role of persistent neural activity during working memory. *Trends in Cognitive Sciences*, *18*(2), 82–89.
- Störmer, V. S., Alvarez, G. A., & Cavanagh, P. (2014). Within-hemifield competition in early visual areas limits the ability to track multiple objects with attention. *The Journal of Neuroscience*, *34*(35), 11526–11533.
- Suchow, J. W., Brady, T. F., Fougner, D., & Alvarez, G. A. (2013). Modeling visual working memory with the MemToolbox. *Journal of Vision*, *13*(10):9, 1–8, <https://doi.org/10.1167/13.10.9>. [PubMed] [Article]
- Talbot, S. A., & Marshall, W. H. (1941). Physiological studies on neural mechanisms of visual localization and discrimination. *American Journal of Ophthalmology*, *24*(11), 1255–1264.
- Theeuwes, J., Atchley, P., & Kramer, A. F. (1998). Attentional control within 3-D space. *Journal of Experimental Psychology: Human Perception and Performance*, *24*(5), 1476–1485.
- Treisman, A., & Zhang, W. (2006). Location and binding in visual working memory. *Memory & Cognition*, *34*(8), 1704–1719.
- 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, USA*, *109*(22), 8780–8785.
- Vogel, E. K., McCollough, A. W., & Machizawa, M. G. (2005, November 24). Neural measures reveal individual differences in controlling access to working memory. *Nature*, *438*(7067), 500–503.
- Wilken, P., & Ma, W. J. (2004). A detection theory account of change detection. *Journal of Vision*, *4*(12):11, 1120–1135, <https://doi.org/10.1167/4.12.11>. [PubMed] [Article]
- Xu, Y., & Nakayama, K. (2007). Visual short-term memory benefit for objects on different 3-D surfaces. *Journal of Experimental Psychology: General*, *136*(4), 653–662.
- Zhang, W., & Luck, S. J. (2008, May 8). Discrete fixed-resolution representations in visual working memory. *Nature*, *453*, 233.

## **Acknowledgement**

Chapter 1, in full, is a reprint of the material as it appears in Journal of Vision, 2019, Chunharas, Chaipat; Rademaker, Rosanne; Sprague, Thomas; Brady, Timothy; Serences, John. The dissertation author was the primary investigator and author of this paper.

## Chapter 2

### Adaptive distortions in visual working memory



## **Summary**

When holding multiple items in memory, representations of individual items are often attracted to, or repelled from, each other. While such distortions in visual working memory are empirically well-established, there is no theoretical framework that explains when and why repulsion or attraction will occur. We propose that repulsion and attraction biases between items reflect different goals of the visual system. We demonstrate that repulsion biases are stronger when two items are highly similar, and these biases grow stronger over time as memory noise increases. Importantly, subjects with better memory performance exhibit stronger repulsion biases – suggesting that biases are not due to a lack of effort but instead confer a functional benefit. By contrast, we find attraction biases when there are more target items than can be remembered with high accuracy (e.g., four instead of two items). Together, these results suggest that repulsion biases reflect a balance between accurate representation and increased inter-item distinctiveness, while attraction biases act to stabilize weak memory signals by capitalizing on information about the entire group of items. Due to resolution and capacity limits, human memory systems are not optimized to accurately represent remembered items; instead, representations are systematically distorted to support specific behavioral goals.

## Results

Memory is a constructive rather than a passive process. For example, people will naturally fill in gaps when recalling a story in an attempt to make the story more coherent (1–3). When people study a list of words, they often falsely recall or recognize associated words that were not on the original list (4, 5), and later report these words as actual memories (6). Similarly, visual memory is not analogous to taking a photo – instead, there are many systematic biases in how visual attributes are remembered after a sensory stimulus is no longer available (7–14).

When people are asked to remember visual information about multiple items, these memories are often 'attracted' to each other - that is, different objects are remembered as more similar than they really were (7, 10, 15–20). It has been proposed that this occurs because object-level representations are imprecise, so these unstable representations are constrained using additional information about the properties of the set of items as a whole (i.e. group-level representation). Thus, attraction biases may be the result of weighting the representation of each individual object towards the “summary” of the set to achieve a more stable memory at the expense of maintaining distinctions between individual items (10, 18).

Interestingly, attraction biases are not ubiquitous. Under some conditions, memories for specific items have been shown to repel each other, being remembered as more different than they really were (9, 21–25). However, far less research has been dedicated to understanding inter-item repulsion biases. Repulsion biases have sometimes been proposed to arise from lateral inhibition, as competition between neurons representing similar feature values may lead to perceptual representations that repel away from each other (26, 27). Despite the importance and pervasiveness of these memory distortions, to date there have been few attempts to understand why memories sometimes are subject to attraction biases and why other times they are subject to repulsion biases.

Here we reconcile these two biases and develop a theory that predicts when each type of bias will occur by considering the potentially adaptive nature of memory distortions (12). We hypothesize that maintaining individuated representations of many memory items is challenging, and relying on group-

level statistics provides an efficient means of retaining at least some information about all items at the expense of precisely representing information about each single item. This strategy is thus adaptive for maximizing accuracy when many items are present. By contrast, when individual items can be maintained without chunking or grouping, and the goal is to distinguish highly similar or noisy representations, a repulsion bias will optimize performance by reducing the confusability of individual representations. This strategy is therefore adaptive when fewer items are present, and when these items are highly confusable.

We tested these predictions in a well-studied domain – visual working memory – where memory representations can be precisely quantified and biases directly observed. Consistent with our theory, we found that when two items are closer together in feature space, or when they're represented with more noise, confusability between items goes up and repulsion biases are stronger. In contrast, we observed attraction biases when individuating items was more difficult due to a higher memory loads, consistent with a strategic choice to sacrifice single-item discriminability in order to remember at least some information about ensemble-level features. Both biases were strong enough that participants were reliably below chance at picking the correct remembered color compared to an appropriately distorted foil color. Collectively, these studies use systematic biases in visual working memory to demonstrate fundamental properties of human memory systems: given task-imposed constraints, attraction or repulsion biases help to optimize mnemonic representations to improve behavioral performance, even though these biases lead to non-veridical memories of features encoded from the sensory environment.

### **Repulsion biases for confusable representations depend on encoding noise.**

In the first experiment (Exp. 1; Figure 1a), we asked participants to remember 2 colors, and we independently manipulated both memory encoding time (50, 150 and 500 ms) and distance in feature space between the two colors ( $20^\circ$ ,  $45^\circ$ ,  $90^\circ$  and  $135^\circ \pm 3^\circ$  on a  $360^\circ$  color-wheel). If two colors repel each other more when they are more confusable (i.e. closer in feature space), more similar colors should result in more repulsion than less similar colors. Furthermore, the color distance that creates maximal repulsion should depend on how precise the representations are, and precision should vary with encoding

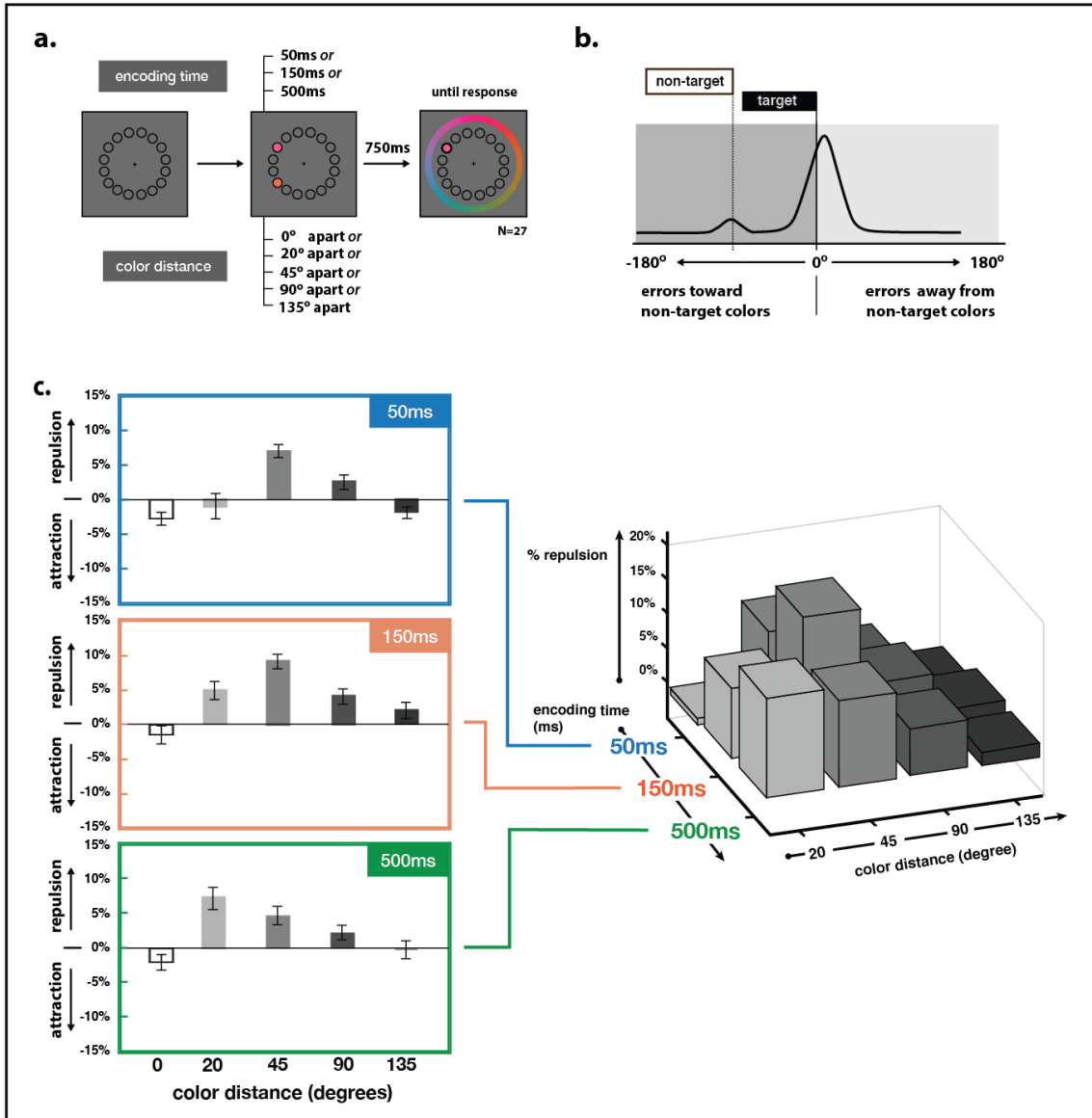
time (i.e. memory should be more precise at longer presentation times). To illustrate: If two colors, 45° apart in color space, are represented with very high precision, they can be easily discriminated without repulsion. However, when increasing noise causes the representations to start overlapping, repulsion could help reduce confusability.

We found that the memory precision changed as a function of encoding time (angular deviations of 26.36°, 25.21°, and 20.62° for encoding times of 50, 150 and 500ms respectively,  $F(2,46) = 65.17$ ,  $p < 0.001$ ) and color distance (circular standard deviations of 17.19°, 21.2°, 25.21°, 29.22° and 29.22° for the color distances of 0°, 20°, 45°, 90° or 135°  $\pm$  3° respectively,  $F(4,92) = 69.49$ ,  $p < 0.001$ ).

To quantify the repulsion effect, we calculated the percentage of responses away from the non-target item (Figure 1b). There were different amounts of repulsion at different encoding times, namely 50.8%, 53.4% and 52.4% for encoding times of 50, 150 and 500ms respectively ( $F(2,46) = 9.185$ ,  $p < 0.001$ ). The percentage of biased responses also differed as a function of distance in color space between the two memory items: The percentage of responses away from the non-target item were 47.8%, 53.8%, 57.0%, 53.0% and 49.4% for color distances of 0°, 20°, 45°, 90° or 135°  $\pm$  3° ( $F(4,92) = 13.14$ ,  $p < 0.001$ ).

Importantly, there was an interaction between encoding time and color distance ( $F(8,184) = 3.777$ ,  $p < 0.001$ ; Figure 1c). For example, the strongest repulsion bias shifted from 45° at the shortest encoding time (50ms) to 20° at the longest encoding time (500ms; Figure 1c). Note how two very similar colors presented at very short encoding times show a decreasing amount of repulsion (with repulsion disappearing altogether when two items were 20° apart and shown for only 50ms). This pattern emerges because under these circumstances people are no longer able to individuate the two items, and they rely on a more gist-like representation. Our initial analyses that were based on a non-parametric quantification of bias were confirmed with an additional analysis based on the bias parameter of a von Mises distribution fit to the data ( $F = 6.98, 8.87, 3.63$ ,  $p < 0.01, < 0.001, < 0.001$  for main effect of encoding time, color distance and interaction between them respectively). Interestingly, memory repulsion does not simply lead to a rigid transformation of two representations in opposing directions. Rather, people also

compress memory representations away from one another, leading to significantly skewed response distributions. Together, these results imply that representations are biased to become more distinctive if their individual representations are more confusable. However, in the limit people ultimately need to individuate items before any repulsion can occur. This means that with shorter encoding times we see maximal repulsion when two items are sufficiently distant in feature space (i.e. at 45° but not 20°). Similarly, it requires very long encoding times to form representations that are sufficiently precise to be confusable at more proximate points in feature space (i.e. 20°), which means that items must be very similar to achieve maximum repulsion.



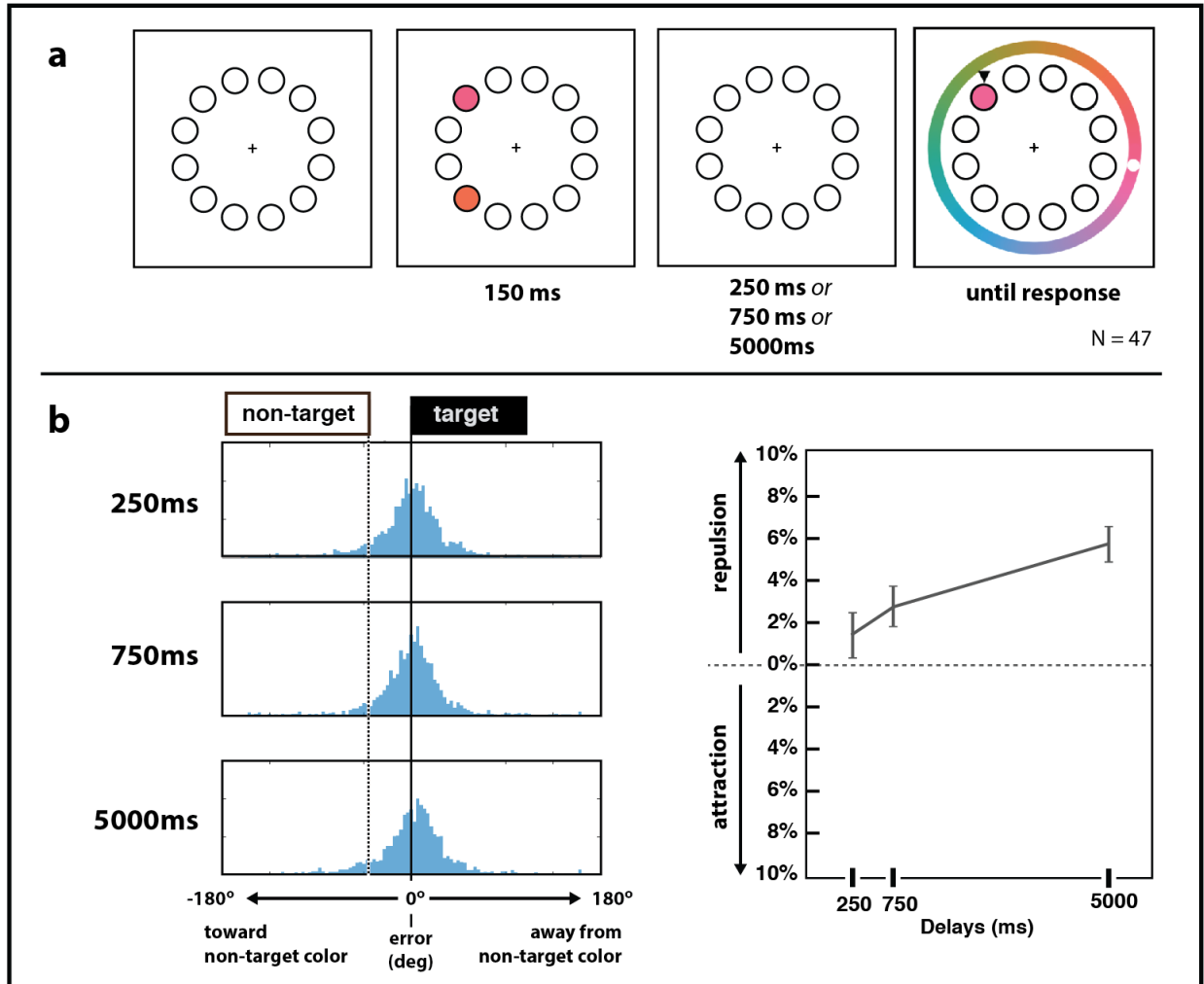
**Figure 2.1: Experimental procedure and results for Experiment 1.** a) In Experiment 1, participants remembered two memory items that were either 0°, 20°, 45°, 90° or 135° apart in CIE  $L^*a^*b^*$  color space, and that were briefly presented for either 50ms, 150ms or 500ms. Participants reported the color of the cued item (indicated by the arrow) by choosing the remembered color on a color-wheel. b) This cartoon demonstrates how, by convention, the error distributions for each subject and condition were presented as if the non-target color was always counter-clockwise from the target color (i.e. on trials with a clockwise non-target color the sign of the error was flipped). Attraction and repulsion biases were operationalized as the difference in the percentage of responses that were toward (dark gray shading) vs. away from (light gray shading) the non-target color, respectively. c.) The 3D bar plot shows repulsion as a function of both encoding time (z-axis) and inter-item distance in color space (x-axis). Repulsion at each level of encoding time is replotted in the three sub-panels to show the within-subject standard error ( $\pm 1$ ) for each condition, and to show the data from trials with a 0° inter-item difference (not shown in the 3D plot) where no repulsion or attraction should exist. Overall, repulsion biases were more prevalent when the two memory colors were more similar. Especially when encoding time increased, and responses become more precise, did the remembered colors need to be very similar to observe maximal repulsion. Error-bars represent  $\pm 1$  within-subject SEM.

### **Repulsion biases for confusable representations grow with longer delays.**

We next asked what happens when memory noise, rather than encoding noise, increases. To manipulate memory noise we compared performance across different memory delay durations, as it has previously been shown that representations become noisier over time (28, 29). We reasoned that if the repulsion bias allows two memories to remain distinct, then the repulsion bias should grow stronger as the memory delay increases. As reported in Experiment 1, this might mean that error distributions become more compressed such that they skew away from the second item. Alternatively, when two representations become increasingly noisy over time, the response may become biased toward the average of the two colors, and thus repulse less, or even attract. Note that in either case, systematic biases would help stabilize behavioral reports when memories become weaker. We tested these predictions in a second experiment (Exp. 2) by manipulating delay duration (250ms, 750ms or 5000ms) while keeping encoding time fixed at 150ms and color distance fixed at 45°. Note that we selected this delay duration and color separation because they yielded the largest repulsion bias in the first experiment.

First, we found that the circular standard deviation was not significantly different across the three memory delays (35.18°, 33.86° and 35.24° for delays of 250ms, 750ms and 5000ms respectively;  $F(2,92) = 0.741$ ,  $p=0.48$ ). Note that previous studies showed that the effect of delay on memory strength is relatively small compared to noise associated with encoding the items and with executing motor responses. Thus, a large number of trials is typically required to yield sufficient power to detect this small effect [28,29] That said, error did significantly increase with longer delays in participants with smaller baseline errors. Finally, inter-item biases might also help stabilize behavioral reports – making it difficult to detect changes in errors even though the memory signal was getting weaker. Indeed, the percentage of responses away from the non-target color grew monotonically with delay duration (51.4%, 52.7%, and 55.6% for delays of 250ms, 750ms and 5000ms respectively;  $F(2,92) = 3.564$ ,  $p=0.032$ ), suggesting larger repulsion biases with increasing delay (Figure 2). This provides further evidence that items are encoded and maintained in a way that increases discriminability. When only two items have to be remembered,

these items are pushed further and further apart with increasing delay, and there is no switch to attraction biases.



**Figure 2.2: Experimental procedure and results for Experiment 2.** Stimulus presentation was similar to Experiment 1, except that encoding time was fixed at 150ms, and the color distance between two memory items was fixed at 45°. The memory delay period was either 250, 750 or 5000 ms. b.) At each delay, error distributions reveal a proportion of responses centered on the non-target, as well as a high number of responses biased away from the non-target (left panels). Despite the presence of non-target responses, the quantified bias still shows a repulsion that grew monotonically stronger as the delay time increased (right panel). Error-bars represent  $\pm 1$  within-subject SEM.



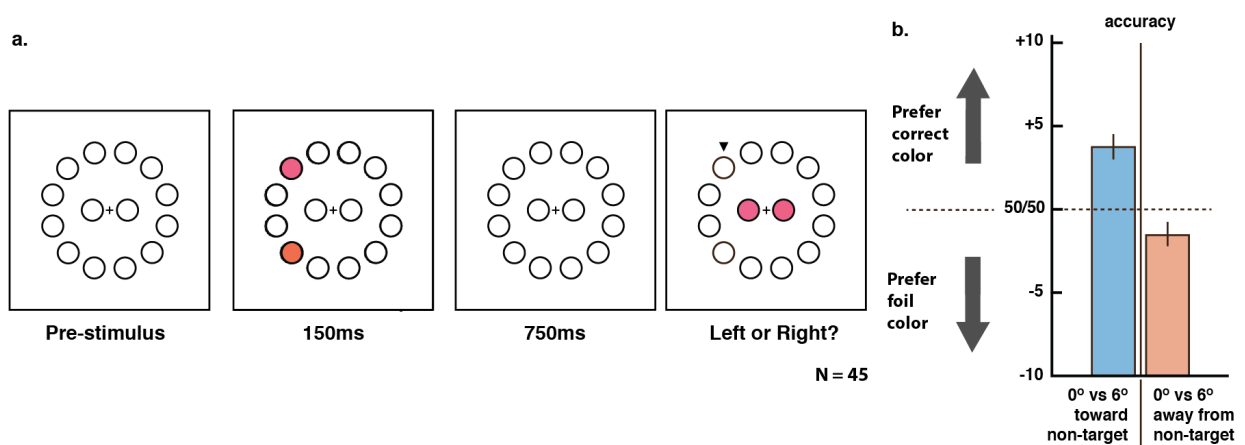
### **Memory distortion vs. response strategy.**

It is possible that the biases we observed in previous experiments did not reflect memory distortions per se, but were instead due to changes in response strategy, such as a strategy to communicate an understanding of the task (e.g., participants wanted to communicate their awareness of the colors being distinct from one another, leading to exaggerated responses on the color-wheel). To address this possibility, we ran a third experiment (Exp. 3) where, instead of responding using a continuous color-wheel, participants performed a two-alternative-forced-choice task comparing the correct (cued) color to an incorrect (distorted foil) color (Figure 3a). We predicted that if memory was truly distorted, and participants actually remembered two items ( $45^\circ$  apart in color-space) as repelled away from each other, a foil color that was distorted *away* from the non-target (by  $6^\circ$ ) would be more easily confused with the correct answer than a foil color that was distorted *toward* the non-target (also by  $6^\circ$ ).

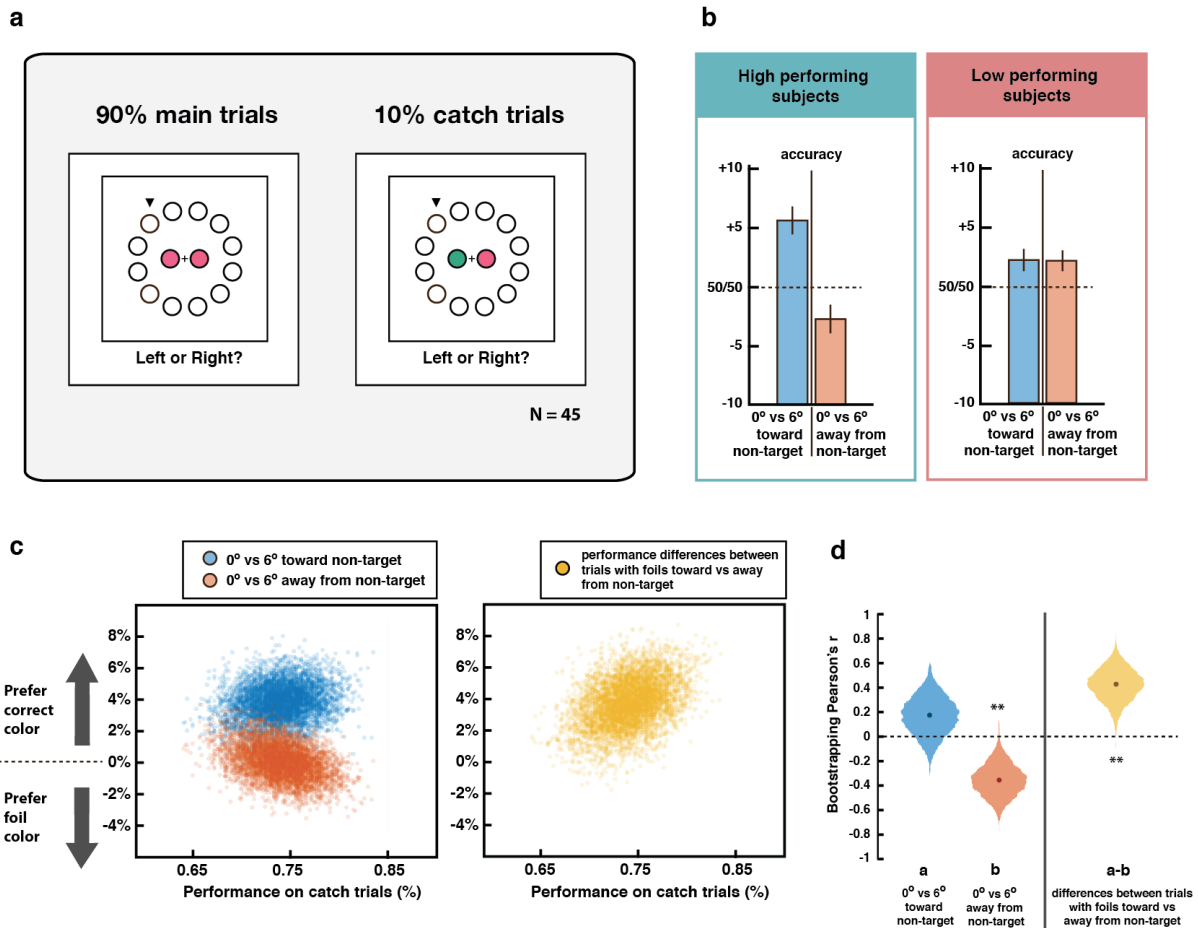
Consistent with this prediction, participants performed better when choosing between the correct color and a foil color that was distorted toward the second (non-target) memory item, and performed worse when choosing between the correct color and a foil color that was distorted away from the second (non-target) memory item ( $t(1,48) = 3.84, p < 0.001$ ; Figure 3b). Participants performed worse when given a foil that was distorted *away* from the non-target color because their memory of the target was also repulsed *away* from the non-target. Thus, they could not tell which of the two response options was correct, and they even exhibited a non-significant trend toward preferring the “distorted-away” (i.e. repulsed) foil color (the correct color was chosen 48.5% of the time;  $t(1,48) = -1.64, p = 0.11$  compared to chance). Participants had fewer problems choosing between the correct color and a foil that was distorted *toward* the second color (the correct color was chosen 53.74% of the time;  $t(1,48) = 3.81, p = 0.004$  compared to chance). This suggests that repulsion biases are truly the result of distorted memory representations, and not of response biases.

We next replicated and extended Experiment 3 with the addition of catch trials in a fourth experiment (Exp. 4). On 10% of trials participants had to choose between a correct color and a foil that was  $180^\circ$  away in color space (Figure 4a). Note that this means there cannot be any attraction or repulsion

biases on catch-trials, as the two colors are on opposing ends of the circular color-space. This added manipulation allowed us to investigate how the repulsion effect related to general task performance (as indexed by performance on catch-trial; Figure 4b, 4c). Because catch-trials were very easy (i.e. the foil color was markedly different from the correct color), we expected performance on these trials to be generally high, and indicative of the level of subjects' task engagement.



**Figure 2.3: Experimental procedure for Experiment 3.** a.) Stimulus presentation was similar to the previous experiments. Encoding time was always 150ms, color distance between two memory items was always fixed at  $45^\circ$ , and the delay interval was always 750ms. Participants reported the color of the cued item (indicated with an arrow) by choosing between two options: the correct color and incorrect foil color that was distorted either  $6^\circ$  toward or away from the non-target color. b.) Participants preferred the correct color to the foil when the foil was distorted toward the non-target color but preferred the foil if it was distorted away from the non-target color. This is the expected result when memories are distorted away from the non-target (i.e. when there is a repulsion bias). Error-bars represent  $\pm 1$  within-subject SEM.



**Figure 2.4: Experimental procedure for Experiment 4.** a.) Stimulus presentation in experiment 4 was identical to that in experiment 3, with the exception that we included 10% of catch-trials. On these catch-trials, participants had to respond by choosing between the correct color and a foil that was 180° away from the correct color. b.) The left panel shows data from the 20 subjects who performed best on catch-trials (they answered correctly more than 75% of the time). The right panel shows data from the 25 subjects who performed worst on catch-trials (they were correct 75% of the time or less). Good subjects showed a pronounced repulsion bias: They were much better able to tell the correct answer from a foil when the foil was distorted towards the second memory item, compared to when a foil was distorted away from the second memory item (compare the blue to the red bar, respectively). By contrast, bad subjects did not show any repulsion bias, and had an equally hard time irrespective of the direction in which the foil was distorted. The did not even perform the main task significantly above chance. c.) Accuracy when choosing between the correct color and a foil color, plotted against general memory performance (indexed by performance on catch trials). Foils could be distorted toward the second non-target memory item (in blue) or distorted away from the non-target item (in red); within-subject differences between these two conditions are shown in yellow in the bottom middle panel. Each dot represents the mean performance on one bootstrapping iteration (of 5000 total iterations) calculated by randomly resampling the data from the 45 participants with replacement. Again, these data demonstrate stronger biases away from the non-target color in participants with better general memory performance. d.) Distribution plots of bootstrapped Pearson's  $r$  for both conditions and the difference between them. \*\* indicates  $p < 0.01$ . Error-bars represent  $\pm 1$  within-subject SEM.

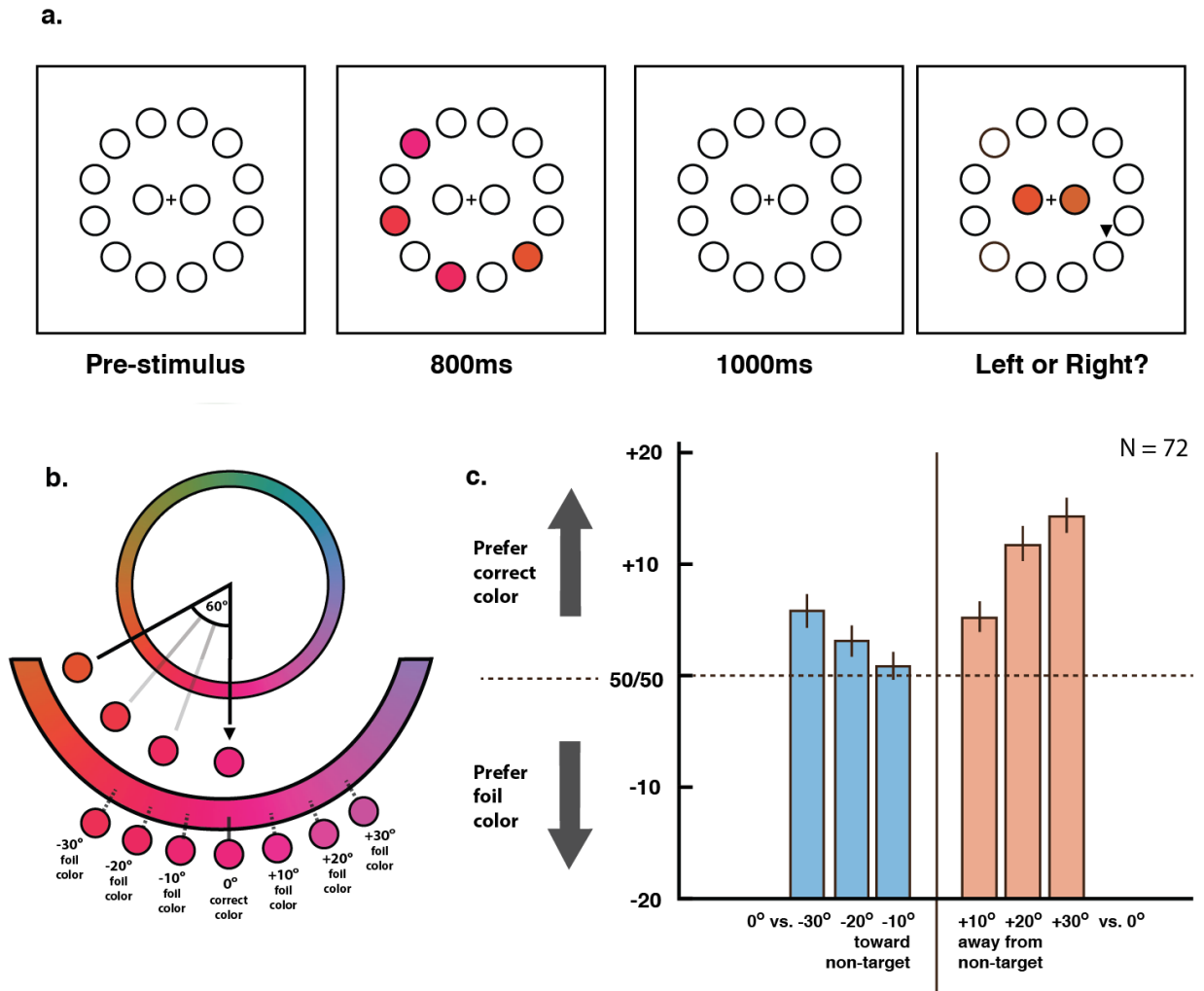
Critically, if the repulsion bias is an adaptive distortion used to improve memory, one should expect the degree of repulsion to positively correlate with memory performance. If biases arise from people not putting in sufficient effort, we would expect it to be negatively correlated with performance. We performed a bootstrapping analysis with 5000 iterations (Figure 4c). On each iteration, data from the 45 subjects were randomly resampled (with replacement) and we calculated their mean accuracy on trials with a foil toward and away from the second (non-target) memory item as well as the difference between them (blue and red dots in Figure 4c and yellow dots in Figure 4d, respectively). We also calculated the average performance on catch-trials across all randomly sampled subjects on each iteration. Better general memory performance resulted in larger repulsion biases, as indicated by a clear separation of the red and blue dot clouds toward the right side of Figure 4c (left) and a positive slope of the yellow dot cloud in Figure 4c (right). We found a moderate correlation between performance on catch-trials ( $180^\circ$  change trials) and the strength of participants' repulsion bias (mean Pearson's  $r$  of 0.39,  $p = 0.009$ , bootstrapping two-sided  $p = 0.012$ , yellow violin plot in Figure 4d). We also evaluated the correlation between performance on catch-trials and accuracy on trials where the foil colors were away vs. toward non-target colors. Performance on catch trials was negatively correlated with accuracy when the foils were shifted away from the non-target color (Pearson's  $r = -0.36$ ,  $p = 0.015$ , bootstrapping two-sided  $p=0.0016$  red violin plot in Figure 4d) but there was a non-significant correlation between catch-trials and were shifted toward the non-target color (Pearson's  $r = 0.18$ ,  $p = 0.250$ , bootstrapping two-sided  $p>0.50$  blue violin plot in Figure 4d). Next, participants were split into low and high performing groups based on a threshold of 75% accuracy on catch-trials (Figure 4b). The high performing group showed the same pattern of results as we observed in Experiment 3, but with a larger magnitude ( $t(19) = 3.68$ ,  $p = 0.0016$ , Cohen's  $d_z$  went from 0.55 in EXP3 to 0.82 in EXP4 high performing group). High performers were good at choosing the correct target color when paired with a foil that was distorted toward the non-target color (above chance mean accuracy of 55.67%;  $t(19)=3.18$ ,  $p=0.005$ ). However, high performers became confused when choosing between the correct color and a foil color that was distorted away from the non-

target color, choosing the foil more often than would be expected by chance (mean accuracy of 47.33%;  $t(19)=2.01$ ,  $p=0.059$ ).

The poor performers did not show differences in accuracy between trials with foils shifted toward or away from the non-target color ( $t(24) = .07$ ,  $p>0.05$ ) and their performance was not significantly above chance when choosing between the correct color and a foil shifted toward or away from the non-target color (mean accuracies of 52.33% and 52.22%, respectively;  $t(24)=1.64$  and  $1.42$ ,  $p = 0.11$  and  $0.17$ ).

### **Attraction vs. repulsion.**

In a final experiment (Exp. 5), we used the same two-alternative-forced-choice paradigm as Experiments 3 and 4, but increased the memory load so that participants now had to remember 4 items on each trial instead of just 2 items. Given well-documented limits on the amount of information that can be retained in WM, remembering 4 items should be quite challenging for the majority of participants (e.g., (30–33)). We hypothesized that when it is challenging to maintain individuated representations of all memory items (e.g., at set size 4) and participants are struggling to achieve reasonable performance levels, then a partial reliance on group-level statistics might provide an efficient means of retaining at least some information about all memory items (e.g., (10)). In this context, when participants are presented with a foil that is distorted *towards* the colors of all the items in the set (Figure 5a), they should be more likely to confuse the foil with the correct (cued) color (i.e. show an attraction bias instead of the repulsion biases reported in our prior experiments).



**Figure 2.5: Experimental procedure and results for Experiment 5.** a.) Participants had to remember a set of four colors (shown in randomly selected locations from the possible set of 12). Participants remembered the items over a 1 second memory delay, and subsequently saw a location cue (triangle) to indicate the memory item to respond to. Two response options were given, and participants chose between the correct (cued) color and a foil color. b.) A set of four colors were selected to lie within  $60^\circ$  of each other in color-space (all separated by steps of  $20^\circ$ ). The target color (to be reported after the delay) was always one of the colors on the edge of the set. The foil color differed by either  $10^\circ$ ,  $20^\circ$  or  $30^\circ$  from the correct (cued) color and could be distorted towards or away from the other four colors in the memory set. c.) Accuracy was lower when subjects had to choose between the correct color and a foil color that was more similar to the target (for example, accuracies were closer to chance for  $\pm 10^\circ$  differences than  $\pm 30^\circ$  differences). Importantly, performance was worse when the foil color was distorted toward the other memory colors in the set (i.e. the blue bars are lower overall than the red bars). Error-bars represent  $\pm 1$  within-subject SEM. Asterisks represent significance levels against chance, with  $** = p < 0.01$  and  $*** = p < 0.001$  after correction for multiple comparisons.

We found that participants were more accurate when the foil colors were more dissimilar, making discrimination easier (53.02%, 57.40% and 60.07% correct when the foils were 10°, 20° and 30° away from the correct color,  $F(1,71) = 13.14$ ,  $p < 0.001$ ; Figure 5b). Importantly, and in line with our prediction, participants were also better at choosing between the correct answer and a foil when the foil color was distorted *away* from the other non-target colors in the set (53.24% vs. 60.42% correct for foil colors distorted *toward* vs. *away*;  $F(1,71) = 15.48$ ,  $p < 0.001$ ; Figure 5b, compare blue and red bars).

## Discussion

Memory, even across short periods of time, is subject to systematic distortions that make mnemonic representations either more separable (repulsion biases) or more similar (attraction biases). While both types of distortion are well documented, there is not an existing theoretical framework that predicts when repulsion or attraction will occur as a function of the type of information being remembered and current task demands. Here, we hypothesized that repulsion biases should arise when highly similar and easily confusable memoranda need to be held in mind in a distinguishable manner. In such situations, representations of individual items in memory should be biased to increase their discriminability. Repulsion will minimize confusion between items, though with some expense to veridical recall. In contrast, we hypothesized that attraction biases will arise when discriminating individual items is less important or challenging (for example, if many items are present). In this case, weaker representations of individual items should be integrated with information about the group-level statistics of all items currently held in memory, making the memory more durable. Thus, an attraction bias could compress the total memory load and enable a reasonable level of performance at the expense of storing precise information about individual items.

Here we completed a set of 5 experiments to test this theoretical framework. In Exp. 1 we found that two memory representations were repelled from each other when the memoranda were highly similar, with the caveat that they were not so similar that they were perceived as the same color. Remembered items required less similarity to elicit repulsion when the stimulus presentation times were shorter and

consequently the encoded memory representations were noisier. Exp. 2 revealed that repulsion biases also grew stronger as representations were remembered over longer delay intervals. Together, these results highlight the role of repulsion in supporting memory for individuated items, particularly when mnemonic representations are weak and prone to confusion. In Exps. 3 and 4, we expanded on these findings and established their generality by using a different experimental paradigm that required participants to perform a two-alternative-forced-choice task to identify a previously remembered feature. This paradigm also revealed robust repulsion effects, and, importantly, participants with better performance showed larger repulsion biases, demonstrating that repulsion serves an adaptive function and does not simply reflect a lack of effort to precisely remember the colors. Finally, in Exp. 5 we increased memory load to four items, which made it harder for participants to maintain information about each of the individual items. We found that memory biases reversed, from repulsion to attraction, demonstrating that participants relied more heavily on summary statistics to compress total memory load when working memory capacity was taxed and individuated representations were difficult to maintain.

### **Mechanisms of memory biases**

Attraction biases can occur both in a more absolute way (e.g., towards particularly salient colors; (34)) or arise from the similarity between items in an individual display (as in the current work). These attraction biases are straightforwardly explained as arising from gist-based or ensemble-based representations, and an optimal combination of these global representations with item specific representations. In particular, many models claim that attraction biases may be the result of weighting the representation of each object towards the “summary” of the set to achieve a more stable memory at the expense of maintaining distinctions between individual items (10) – an adaptive strategy that serves to minimize error (19).

Repulsion biases have traditionally been more difficult to understand. Previous studies have shown that repulsion biases occur when two items are task-relevant and proximal in features space [9,21]. However, the benefits repulsion biases are still unclear. Here, we propose that repulsion biases serve to



maximize distinctiveness between items, when item representations are strong. Any factor that affects distinctiveness in memory should thus impact the degree of repulsion biases (e.g. encoding time, feature similarity, memory delay, number of items). Interestingly, previous work has frequently found repulsion not only between items, as in the current work, but also in absolute terms as well. For example, when asked to remember an orientation that is near, but not quite at, vertical, people will systematically report the orientation as further from vertical than it really was (35, 36). One framework that has been useful to understand these absolute biases is to dissociate the physical space of the stimuli (e.g., absolute orientation) from the psychological representation of the stimuli (e.g. people may over-weight certain values in a systematic manner). A clear example of a warped psychological space is the massive overrepresentation of vertical and horizontal orientations, perhaps to efficiently code environmental regularities (37, 38). Accounting for this selective over-representation of certain stimulus values in psychological space can explain biases like repulsion from cardinal axes, and the reason why these biases tend to arise in parts of stimulus space where discrimination thresholds are lowest (e.g., the most overrepresented stimulus values) (37, 39).

This conception of psychological space is designed to address long-term biases that are likely crystalized in the neural architecture of the visual system, whereas the biases we examine in the current work are more dynamic and rapidly adapt to task demands. Despite the apparent disconnect, there are reasons to believe that a mechanism such as the warping psychological space may be at play in both stable long-term phenomena and in more dynamic short-term regimes. For instance, spatial judgments are distorted by top-down factors such that there is repulsion bias away from currently attended locations (25). Attention, which leads to well-documented changes in visual sensitivity (i.e., lower discrimination thresholds, see (40)), may therefore serve as a quick and adaptive way to systematically bias perception and memory on demand. After all, biases typically manifest when discrimination thresholds are low across a variety of visual features such as orientation, motion direction, spatial frequency, and visual speed (see [37] for a summary). In sum, conceptions of psychological space, and how it is distorted when

a particular set of stimuli is over-represented, may be a useful framework for considering biases at all possible time scales (see also (41), for details on the widely applicable utility of this concept).

What might be the neural substrates of biased representations? When a task requires focal attention to a small set of items to remember – as is the case in paradigms that create repulsion bias – the discriminability of the relevant items can be improved by biasing responses in early visual cortex to maximize the discriminability of their corresponding neural representations. For example, attention to highly similar features, akin to remembering highly similar colors in Exps. 1-4, has been shown to modulate neurons tuned just away from the attended features. This ‘off-target’ gain can improve performance because neurons tuned away from the attended features undergo the largest change in firing rates because the two features fall along the steepest part of their Gaussian tuning curves. In turn, this off-target gain gives rise to systematic biases in behavioral reports such that people see stimuli as repulsed from the actual feature values (42–44). Such repulsion would be expected if the off-target gain happening in early visual cortex was interpreted as a veridical representation of the world at higher stages of processing. While previous work in this domain has focused on selective attention to continuously present stimuli, a similar type of modulation in the domain of WM might give rise to repulsive biases in mnemonic representations. While speculative, this type of adaptive neural modulation may map onto the psychological space framework, such that changes in the discriminability of stimuli in early visual cortex – either due to a lifetime of experience or to dynamic changes in the focus of attention – lead to a warping of perception and memory.

## **Methods**

### **Experiment 1**

#### *Participants*

Twenty-four healthy volunteers (15 female, mean age of  $19.75 \pm 1.52$ ) from the University of California San Diego (UCSD) community participated in the experiment. All procedures were approved by the UCSD Institutional Research Board and all participants provided written informed consent, and

reported normal or corrected-to-normal vision without color-blindness. Participants were naïve to the purpose of the study and received partial course credit for their time.

### *Stimuli*

Stimuli were rendered on a NEC-Mitsubishi MultiSync FP912SB CRT monitor with a 60-Hz refresh rate and a screen size of 17.4" x 17.4". Stimuli were generated on a PC running Ubuntu (v16.04) using MATLAB and the Psychophysics toolbox (45, 46). Participants were instructed to maintain fixation throughout, aided by a white central fixation dot (0.25° diameter) presented on a dark-gray background of 2.37 cd/m<sup>2</sup>. Memory items were colors selected from a subset of CIE color space ( $L = 70, a = 20, b = 38$ , radius = 60). Two memory targets were 0°, 20°, 45°, 90° or 135° apart in the circular color space (with  $\pm 3^\circ$  random jitter) and always presented at a spatial separation of 3° of visual angle. Sixteen placeholders were positioned around the fixation point (3.5° from fixation, 0.6° in diameter, white rim with thickness of 0.02°). The location of the memory targets were selected at random except that they were always presented in the same hemifield to maximize inter-item competitions (47–49).

### *Procedure*

On each trial, two colored stimuli were presented for either 50ms, 150ms or 500ms and participants had to remember the colors as precisely as possible. After a 750ms delay, one of the two colors was probed via a spatial cue (the rim of placeholder in one location got thicker). Along with the spatial probe, a color-wheel (with 3° radius, 0.3° wide) was presented around fixation, and a crosshair appeared at the fixation point. Participants used the mouse to move the crosshair to the hue on the color-wheel that most closely resembled the remembered color at the probed location. The next trial began ~1s after participants clicked the mouse and this procedure was repeated 96 times per experimental condition (i.e. a total of 1440 trials per participant). Presentation of the 5 different color distances and 3 different encoding times was fully counterbalanced.

### *Analyses*

We generated a distribution of errors for each participant by calculating the difference between the cued target color and the reported color (reported° – target°) on each trial. To clearly visualize the

shape of the error distribution and its relationship to the non-target color, we flipped the sign of the error on trials where the non-target color was clockwise of the target color. Thus, all data presented here are sorted such that the non-target color was always counter-clockwise to the cued target. The precision of subject's responses was quantified by calculating the circular standard deviation of the error distribution. Biases in subjects report were quantified by computing the proportion of responses on the clockwise side of the error distribution. If no biases exist, both sides of the error distribution (centered on a  $0^\circ$  error) should contain roughly half of the errors made by participants. However, because the non-target is always on the counter-clockwise side of the distribution, we expect this bias metric to be  $>0.5$  if there is repulsion away from the non-target, and  $<0.5$  if there is attraction toward the non-target. Note that this metric reflects relative repulsion/attraction biases rather than an absolute metric, since any swap errors (where the target and foil colors are confused, and a subject mistakenly reports the foil) would be counted as "attraction". Thus, this metric is conservative to the extent that swap errors inflate attraction biases and underestimate repulsion biases. To benchmark our model-free metrics of memory precision and bias, we also fit a von Mises (circular analogue of a normal distribution) to our error distributions using 2 parameters: concentration ( $k$ ) and bias ( $\mu$ ). We used repeated-measures analysis of variance to evaluate the impact of encoding time and color similarity on both the model-free (percentage of responses away from the non-target) and estimated (von Mises) model parameters.

## **Experiment 2**

### *Participants*

Forty-seven participants were recruited using Amazon Mechanical Turk. All participants were from the United States, provided their informed consent (approved by UCSD Institutional Review Board), and were paid \$2 for approximately 10 min of their time.

### *Stimuli*

All stimuli were drawn on a  $400 \times 400$  pixels white background with a black border around it (1 pixel wide). The fixation point was in the middle of the canvas, and 12 small circular placeholders were

shown around fixation, at a distance of 120 pixels. Each placeholder had a radius of 20 pixels, and the inter-placeholder distances were 62 pixels. Placeholders were positioned such that six of them were on the left, and the other six were on the right side of fixation. Memory colors were selected from a circular subset of CIE color space ( $L = 70, a = 20, b = 38, \text{radius} = 60$ ), such that the two colors selected on each trial were always  $45^\circ$  apart in color space. The location probe was a small equilateral black triangle, 20 pixels wide and 20 pixels tall.

### *Procedure*

On each trial, two stimuli were presented for 150 ms and participants had to remember the color of each stimulus for a 250 ms, 750 ms, or 5000 ms delay period. After the delay, one of the two colors was probed by presenting a small arrow above one of the placeholders, pointing out the target location. A color-wheel (170 pixels radius, 10 pixels wide) and a crosshair appeared together with the location probe. As in Exp. 1, participants used the mouse to move the crosshair from its initial location at fixation, to the hue on the color-wheel that most closely resembled the color remembered at the probed location. The next trial began 2s after participants clicked the mouse to indicate their answer, and this procedure was repeated 36 times for each of the 3 experimental conditions (108 trials per participant in total).

### **Experiment 3-5**

A new set of naïve participants were recruited from Amazon Mechanical Turk for each of the three experiments described here (experiments 3–5). For experiments 3 and 4 we recruited 45 participants each, for experiment 5 we recruited 72 people, due to the increased difficulty of the task associated with the higher set size (thus requiring more power). The display and trial structure were the same as in previous experiments. Items could appear at 12 possible locations that were equidistant from fixation. Encoding time (150ms in experiments 3 and 4; 800ms in experiment 5) and delay time (750ms in experiments 3 and 4; 1000ms in experiment 5) were fixed, and participants remembered either 2 memory items that differed by  $45^\circ$  in color space (experiments 3 and 4) or 4 memory items (experiment 5). In experiment 5, the four colors were chosen from CIE color space ( $L = 70, a = 20, b = 38, \text{radius} = 60$ )

within  $60^\circ$  from each other on the color wheel, and equally spaced (i.e. shortest color distance of each item was  $20^\circ$ ) (Figure 5A). The memory target probed at the end of the delay was always one of the colors at the edge of the set (and thus furthest away from the mean of the color set). The spatial location at which each of the four colors was shown was entirely random. During the response, instead of using a color wheel to recall color in a continuous manner, this set of three experiments relied on a two-alternative-forced-choice task. After the delay, a location cue (arrow) indicated which item was the memory target, and two color-choices appeared on both sides of the fixation point. In experiments 3 and 4, one of the two choices was always the target color, while the other was a foil whose color differed by  $6^\circ$  from the correct target. This  $6^\circ$  distortion was either toward (50% of the trials) or away (50% of the trials) from the second non-target color. In experiment 5 the correct color was also always included as a response option, while the foil color differed by either  $10^\circ$ ,  $20^\circ$ , or  $30^\circ$  from the correct target color. A foil color could be either *toward* the colors of the other 4 memory items (note how a  $-20^\circ$  foil is identical to one of the other colors in the display, and a  $-30^\circ$  foil is exactly the mean of all 4 colors), or it could be *away* from the other four colors. During the response phase of all three experiments described here, the positions of the correct and foil colors were completely randomized. Participants had to press “z” or “m” to select the choice presented on the left or right of fixation before proceeding to the next trial.

## **Acknowledgement**

Chapter 2, in part, is currently being prepared for submission for publication of the material. Chunharas, Chaipat; Rademaker, Rosanne; Brady, Timothy; Serences, John. The dissertation author was the primary investigator and author of this material.

## References

1. Bartlett FC (1932) Some Experiments on the Reproduction of Folk-Stories. *Folklore* 31(1):30–47.
2. Roediger, H. L., & McDermott, K. B. (1995) Creating false memories: Remembering words not presented in Lists. *J Exp Psychol Learn Mem Cogn* 21(4):803.
3. Loftus EF (2005) Planting misinformation in the human mind: a 30-year investigation of the malleability of memory. *Learn Mem* 12(4):361–366.
4. Deese J (1959) On the prediction of occurrence of particular verbal intrusions in immediate recall. *J Exp Psychol* 58(1):17–22.
5. Underwood BJ, Ekstrand BR, Keppel G (1965) An analysis of intralist similarity in verbal learning with experiments on conceptual similarity. *Journal of Verbal Learning and Verbal Behavior* 4(6):447–462.
6. Schooler JW, Foster RA, Loftus EF (1988) Some deleterious consequences of the act of recollection. *Mem Cognit* 16(3):243–251.
7. Huang J, Sekuler R (2010) Distortions in recall from visual memory: two classes of attractors at work. *J Vis* 10(2):24.1–27.
8. Alvarez GA (2011) Representing multiple objects as an ensemble enhances visual cognition. *Trends Cogn Sci* 15(3):122–131.
9. Rademaker RL, Bloem IM, De Weerd P, Sack AT (2015) The impact of interference on short-term memory for visual orientation. *J Exp Psychol Hum Percept Perform* 41(6):1650–1665.
10. Brady TF, Alvarez GA (2011) Hierarchical encoding in visual working memory: ensemble statistics bias memory for individual items. *Psychol Sci* 22(3):384–392.
11. Fischer J, Whitney D (2014) Serial dependence in visual perception. *Nat Neurosci* 17:738.
12. Schacter DL, Guerin SA, St. Jacques PL (2011) Memory distortion: an adaptive perspective. *Trends Cogn Sci* 15(10):467–474.
13. Koutstaal W, Verfaellie M, Schacter DL (2001) Recognizing identical versus similar categorically related common objects: further evidence for degraded gist representations in amnesia. *Neuropsychology* 15(2):268–289.
14. Bar M (2004) Visual objects in context. *Nat Rev Neurosci* 5(8):617–629.
15. Dubé C, Zhou F, Kahana MJ, Sekuler R (2014) Similarity-based distortion of visual short-term memory is due to perceptual averaging. *Vision Res* 96:8–16.
16. Freyd JJ, Johnson JQ (1987) Probing the time course of representational momentum. *J Exp Psychol Learn Mem Cogn* 13(2):259–268.
17. Hemmer P, Steyvers M (2009) A Bayesian account of reconstructive memory. *Top Cogn Sci* 1(1):189–202.

18. Huttenlocher J, Hedges LV, Duncan S (1991) Categories and particulars: prototype effects in estimating spatial location. *Psychol Rev* 98(3):352–376.
19. Huttenlocher J, Hedges LV, Vevea JL (2000) Why do categories affect stimulus judgment? *J Exp Psychol Gen* 129(2):220–241.
20. Spencer JP, Hund AM (2002) Prototypes and particulars: geometric and experience-dependent spatial categories. *J Exp Psychol Gen* 131(1):16–37.
21. Bae, G. Y., & Luck, S. J. (2017) Interactions between visual working memory representations. *Atten Percept Psychophys* 79(8):2376–2395.
22. Golomb JD (2015) Divided spatial attention and feature-mixing errors. *Atten Percept Psychophys* 77(8):2562–2569.
23. Rauber HJ, Treue S (1998) Reference repulsion when judging the direction of visual motion. *Perception* 27(4):393–402.
24. O’Toole B, Wenderoth P (1977) The tilt illusion: repulsion and attraction effects in the oblique meridian. *Vision Res* 17(3):367–374.
25. Suzuki S, Cavanagh P (1997) Focused attention distorts visual space: an attentional repulsion effect. *J Exp Psychol Hum Percept Perform* 23(2):443–463.
26. Johnson JS, Spencer JP, Luck SJ, Schöner G (2009) A dynamic neural field model of visual working memory and change detection. *Psychol Sci* 20(5):568–577.
27. Wei Z, Wang X-J, Wang D-H (2012) From distributed resources to limited slots in multiple-item working memory: a spiking network model with normalization. *J Neurosci* 32(33):11228–11240.
28. Shin H, Zou Q, Ma WJ (2017) The effects of delay duration on visual working memory for orientation. *J Vis* 17(14):10.
29. Rademaker RL, Park YE, Sack AT (2018) Evidence of gradual loss of precision for simple features and complex objects in visual working memory. *Journal of Experimental Psychology* 44(6):925–940.
30. Luck SJ, Vogel EK (1997) The capacity of visual working memory for features and conjunctions. *Nature* 390(6657):279–281.
31. Ma WJ, Husain M, Bays PM (2014) Changing concepts of working memory. *Nat Neurosci* 17(3):347–356.
32. Bays PM, Catalao RFG, Husain M (2009) The precision of visual working memory is set by allocation of a shared resource. *J Vis* 9(10):7.1–11.
33. Bays PM (2015) Spikes not slots: noise in neural populations limits working memory. *Trends Cogn Sci* 19(8):431–438.
34. Bae G-Y, Olkkonen M, Allred SR, Flombaum JI (2015) Why some colors appear more memorable than others: A model combining categories and particulars in color working memory. *J Exp Psychol Gen* 144(4):744–763.



35. Jastrow J (1892) Studies from the University of Wisconsin: on the judgment of angles and positions of lines. *Am J Psychol* 5(2):214–248.
36. Smith SL (1962) Angular estimation. *J Appl Psychol* 46(4):240–246.
37. Wei, X. X., & Stocker, A. A. (2015) A Bayesian observer model constrained by efficient coding can explain “anti-Bayesian” percepts. *Nature neuroscience* 18(10):1509–1517.
38. Girshick AR, Landy MS, Simoncelli EP (2011) Cardinal rules: visual orientation perception reflects knowledge of environmental statistics. *Nat Neurosci* 14(7):926–932.
39. Wei, X. X., & Stocker, A. A. (2017) Lawful relation between perceptual bias and discriminability. *Proceedings of the National Academy of Sciences* 114(38):10244–10249.
40. Carrasco M (2011) Visual attention: the past 25 years. *Vision Res* 51(13):1484–1525.
41. Schurgin M.W., Wixted J.T., Brady T.F. (2018) Psychological Scaling Reveals a Single Parameter Framework For Visual Working Memory. *BioRxiv*. Available at: <https://www.biorxiv.org/content/biorxiv/early/2018/05/18/325472.full.pdf>.
42. Navalpakkam V, Itti L (2007) Search goal tunes visual features optimally. *Neuron* 53(4):605–617.
43. Scolari M, Serences JT (2009) Adaptive allocation of attentional gain. *J Neurosci* 29(38):11933–11942.
44. Jazayeri M, Movshon JA (2007) A new perceptual illusion reveals mechanisms of sensory decoding. *Nature* 446(7138):912–915.
45. Brainard DH (1997) The Psychophysics Toolbox. *Spat Vis* 10(4):433–436.
46. Pelli DG (1997) The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spat Vis* 10(4):437–442.
47. Cohen MA, Rhee JY, Alvarez GA (2016) Limits on perceptual encoding can be predicted from known receptive field properties of human visual cortex. *J Exp Psychol Hum Percept Perform* 42(1):67–77.
48. Störmer VS, Alvarez GA, Cavanagh P (2014) Within-hemifield competition in early visual areas limits the ability to track multiple objects with attention. *J Neurosci* 34(35):11526–11533.
49. Alvarez GA, Cavanagh P (2005) Independent resources for attentional tracking in the left and right visual hemifields. *Psychol Sci* 16(8):637–643.

## Chapter 3

Selective amplification of salient features of visual memories during early memory consolidation

## **Summary**

When people need to remember a whole set of words or images, they tend to remember any particular item as more closely resembling the gist of the entire set than it really was. Here we show that when participants are asked to remember only a few items -- and so maintaining the distinctiveness of this item in memory is of particular importance -- memory for that item is distorted in the opposite direction, amplifying its salient features. In a sequence of 4 experiments, we asked participants to remember the aspect ratio and size of a rectangle and draw it after various delays. Participants reliably exaggerated its distinctive feature in every experiment. This distortion occurred not just at initial encoding but also during memory consolidation and persisted for several hours. Thus, when remembering only a few items, memory amplifies the distinctive features of these items, a form of adaptive memory distortion.

## **Introduction**

Due to limitations inherent in any biological system, our memories of recent events are not as vivid, authentic and reliable as they seem to us. Instead of recording every bit of information like a video camcorder, our memories are constructed – they are abbreviated, edited and retrieved in a manner consistent with pre-existing schemata (Bartlett, 1932). Such memory distortion is not random, but instead a property of how our memory system works (Schacter, Guerin, & St. Jacques, 2011). One commonly found example of memory distortion is the systematic distortion of memory towards the gist information (Huttenlocher, Hedges, & Duncan, 1991; Huttenlocher, Hedges, & Vevea, 2000). Such distortions have been found in working memory, long-term memory, in the relationship between episodic and semantic memory, and for nearly all kinds of stimuli (Alvarez, 2011; Brady & Alvarez, 2011; Dubé, Zhou, Kahana, & Sekuler, 2014; Freyd & Johnson, 1987; Hemmer & Steyvers, 2009; Huang & Sekuler, 2010; Huttenlocher et al., 1991, 2000; Spencer & Hund, 2002). In all of these cases, when people need to remember the properties of an object and report them later, the responses tend to be shifted toward the prototype or the center of the relevant category. Computational models show that this helps us to minimize error, as a response toward the mean is less likely to be incorrect compared with a response away from the mean (Brady & Alvarez, 2011; Huttenlocher et al., 1991, 2000), and this property of memory may also promote memory stability.

However, systematic biases toward the center of a category (i.e., attraction bias) are not always useful. For example, if we need to remember one particular object among other similar things (e.g. a particular dog at the dog park), attraction toward the center of the category would reduce distinctiveness and create more confusions. Under those circumstances, it may instead be optimal for memory to be biased away from the gist: maximizing the distinctive feature (i.e., contrast). This should result in memory holding a caricature version of that object (e.g., one more different from the center of the category than the actual item) instead of blending it together with the other similar objects. While attenuation of distinctive features (blending) promotes generalization, amplification of distinctive features

may promote specificity. There is some evidence consistent with this idea, but in general, existing work supposes that that people's underlying representation of the stimulus features is unbiased and veridical. For example, previous work has shown that when presented with a caricature, people are more likely to recognize it and do so more quickly (Lee, Byatt, & Rhodes, 2000; Mauro & Kubovy, 1992). However, this is generally interpreted as showing that, while the underlying representation is accurate and un-caricatured, the matching process needed for recognition benefits from more extreme stimuli (which resemble other faces even less than a normal stimulus, giving a competitive advantage to the correct stimulus; e.g., Tanaka, 1996). Similarly, there is a significant literature on pattern separation in the hippocampus, particularly in the dentate gyrus; however, this literature largely assumes that only the cues needed to retrieve the object, not the actual object features, are made more distinctive (e.g., Yassa & Stark, 2011).

Thus it remains largely unknown whether the distinctive features of the object themselves are amplified or biased away from the category or prototype to enhance their distinctiveness in memory. To address this, we asked participants to remember simple objects with clear, distinctive features. We hypothesized that those features should be amplified during subsequent testing. To ensure that it was the representation itself that was affected and not the matching process, we use participants' drawing to measure their memory (a visual free recall). As our main test case, we ask about memory for a very simple stimulus: a rectangle. We hypothesized that for a rectangle, the intuitive comparison point is a square, and thus the aspect ratio between the 2 sides of the rectangle should become more pronounced in memory, i.e., participants should remember a more extremely rectangular rectangle. Our results showed evidence of this amplification of distinctive feature effect, as the ratio of memorized rectangles was larger after delays even when the rectangular ratios of the memoranda were initially larger than individual's prototypical rectangle. Interestingly, the effect was present even when a rectangle was the only item that subjects needed to remember. In follow-ups, we found that the amplification was larger at 7 minutes compared to 2 minutes delay – indicating that the amplification effect continues to strengthen with increasing delay. Thus, contrary to popular wisdom, memory formation involves more than silent

incubation, memories are changed to make them more distinctive as they are consolidated, a process we term *distinctiveness exaggeration* or *caricaturization* of memories.

## **Experiment 1**

We first investigated whether the memory for the ratio of the sides of a rectangle was amplified by asking participants to remember an image with a rectangle and later draw it from memory. Since a significant previous literature suggests that rather than amplification we should expect the ratio of the drawn rectangle to become closer to the prototypical rectangle e.g. (Hemmer & Steyvers, 2009; Huttenlocher et al., 2000) we first elicited participants' prototypical rectangle from them and we designed a one-trial recall task where the stimuli to be remembered were already *more rectangular* than each participants' own rectangle prototype. Thus, if the ratio of the sides of the rectangle is amplified even further, such amplification cannot be explained by attraction towards the prototypical rectangular ratio.

## **Methods**

### *Participants*

We based the effect size on a similar drawing task of the well-known boundary extension effect (minimum number of participants of 6 with previous reported Cohen's  $d$  of  $\sim 1.8$ ,  $\alpha = 0.05$ , power = 0.80) (Hubbard, Hutchison, & Courtney, 2010; Intraub & Richardson, 1989). Twelve healthy volunteers (7 female, mean age of  $20.9 \pm 0.6$  S.E.M.) from the University of California San Diego (UCSD) community participated in the experiment. All procedures were approved by the UCSD Institutional Research Board. All participants reported normal or corrected-to-normal vision without color-blindness and provided written informed consent. Participants were naïve to the purpose of the study and received course credits for their time. Three participants failed to show up to both sessions and so did not produce data.

### *Stimuli*

Since each person may have a different prototypical rectangular ratio, we first estimated each individual's prototypical rectangular ratio. To do so, we asked each participant to draw a rectangle -- the first rectangle that comes to mind. Then, we measured all four sides of drawn rectangles and computed the ratio by dividing the average length of two long sides by the average of two short sides. The computed ratios were used to make a new, exaggerated rectangle (1.2 times more rectangular than the ratio they drew) that each participant had to remember in the main experiment. To distract subjects from the main goal of the experiment, we included multiple bits of task-irrelevant information. Specifically, the memoranda consisted of two geometric shapes - a specific rectangle and a square (with the same total area). They had different colors and positions such that the lengths of the shapes could not be directly compared (Figure 1). Participants were instructed to remember all of this information.

### *Procedure*

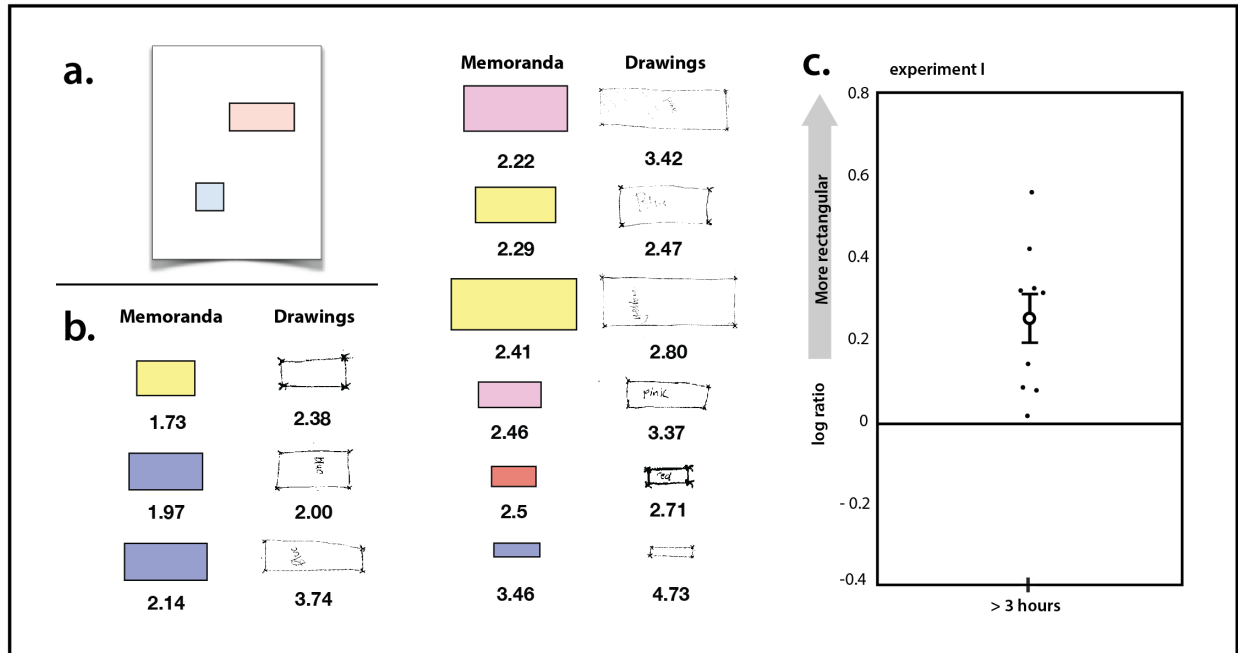
Two weeks after subjects initially drew their prototypical rectangle, we gave them a specific memorandum. This was specifically created for each person and instructed them to "remember every detail on the paper as precisely as you can". After a 3 hour delay, each participant received a blank sheet of paper and was instructed to "draw everything that you saw earlier as precisely as possible" (this was designed to avoid using labeling words such as "shape" or "rectangle" that might explicitly evoke the concept of rectangularity).

### *Analysis*

We measured the ratio of the drawn rectangle by dividing the average length of the long sides by the average of short sides. The degree of amplification was the log of the ratio between the response ratio and the ratio of the 1.2x-exaggerated rectangle in the memorandum. Thus, the number was zero if there were no bias in any direction; positive and negative values if the response rectangle was more and less extreme than the one they had to remember respectively. The log scale means that 2x and 0.5x become the same magnitude with a different sign (0.69 and -0.69 respectively in this example). A one-sample two-sided t-test was performed to demonstrate statistical differences.

## Results

All participants drew skinnier rectangles (9 out of 9, Figure 1C.,  $t(8) = 4.156, p=0.003$ , Cohen's  $d$  of 1.39) than the original memoranda. This distortion was not toward the prototypical rectangular ratio as the rectangle in the memoranda was already more rectangular than the prototype. Thus, we find distinctiveness exaggeration - the rectangle is remembered as more rectangular than it was.



**Figure 3.1: Data from Experiment 1.** Data from Experiment 1. a.) Example of an image generated for a particular participant; Note that we included a square as a distractor. b.) Each pair shows the rectangle a given participant studied (in color) and then the rectangle drawn by that same participant when their memory was subsequently probed. The numbers are the ratios of the width to the height. All subjects drew a more amplified rectangle after >3 hours delay. c) The degree of amplification as quantified by the log ratio of the response drawing ratio vs. study item ratio was significantly higher than 0. The error bar is +/- 1 standard errors of the mean (S.E.M.) and each dot represents the data from one participant.

## Experiment 2

Our stimulus in Experiment 2 consisted not only of the rectangle but also of a square shape that served as a distractor. Thus, in Experiment 2 we examined whether the effect in Experiment 1 was driven by the presence of the square. In particular, we asked whether the ratio of the rectangle would still be exaggerated even without the square. In addition, we asked whether this distinctiveness exaggeration



effect was a result of initial encoding, or whether it changed in memory. To answer this, we tested at different time delays, ranging from 2 minutes to 7 days.

## **Methods**

### *Participants*

We predicted that the exaggeration effect will be substantially smaller when there was only one object to remember. Thus, we set the effect size at 1.0 (a minimum number of participants of 16,  $\alpha = 0.05$ , power = 0.80). Twenty-one healthy volunteers (14 female, mean age of  $20.9 \pm 0.6$  S.E.M.) from the University of California San Diego (UCSD) community participated in the experiment. All procedures were approved by the UCSD Institutional Research Board. All participants reported normal or corrected-to-normal vision without color-blindness and provided written informed consent. Participants were naïve to the purpose of the study and received course credit for their time. Four subjects failed to show up to both study sessions and excluded from the experiment.

### *Stimuli*

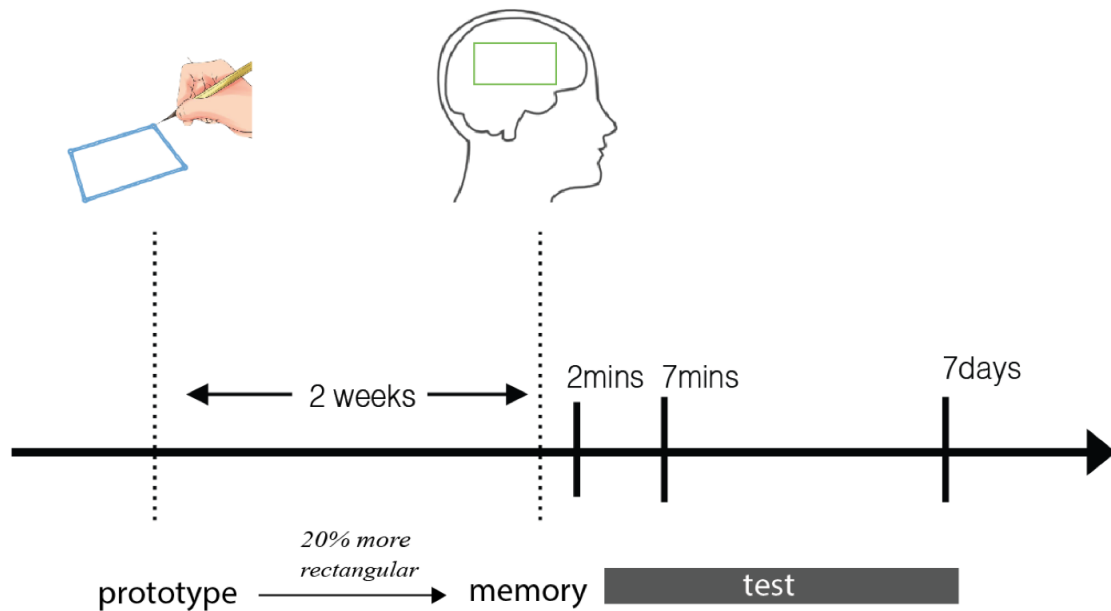
We repeated the stimulus preparation step from Experiment 1; in particular, we acquired each individual's prototypical rectangles a few weeks before the memory experiment and created exaggerated rectangles for each participant. In Experiment 2, however, we now presented these exaggerated rectangles without a square. In particular, the rectangle was now in the middle of the page and slightly tilted so that participants needed to remember not just the rectangle's aspect ratio but also the orientation of the rectangle as an additional feature.

### *Procedure*

The same instructions were given to the participants. After encoding the memoranda, participants then performed a distractor task (passive listening to a story). Unbeknownst to the subjects, they then had to reproduce the rectangle at both 2 and 7 minutes after initial memory acquisition (Figure 2). We also repeated the test after 7 days to measure the effect of a longer delay.

### *Analysis*

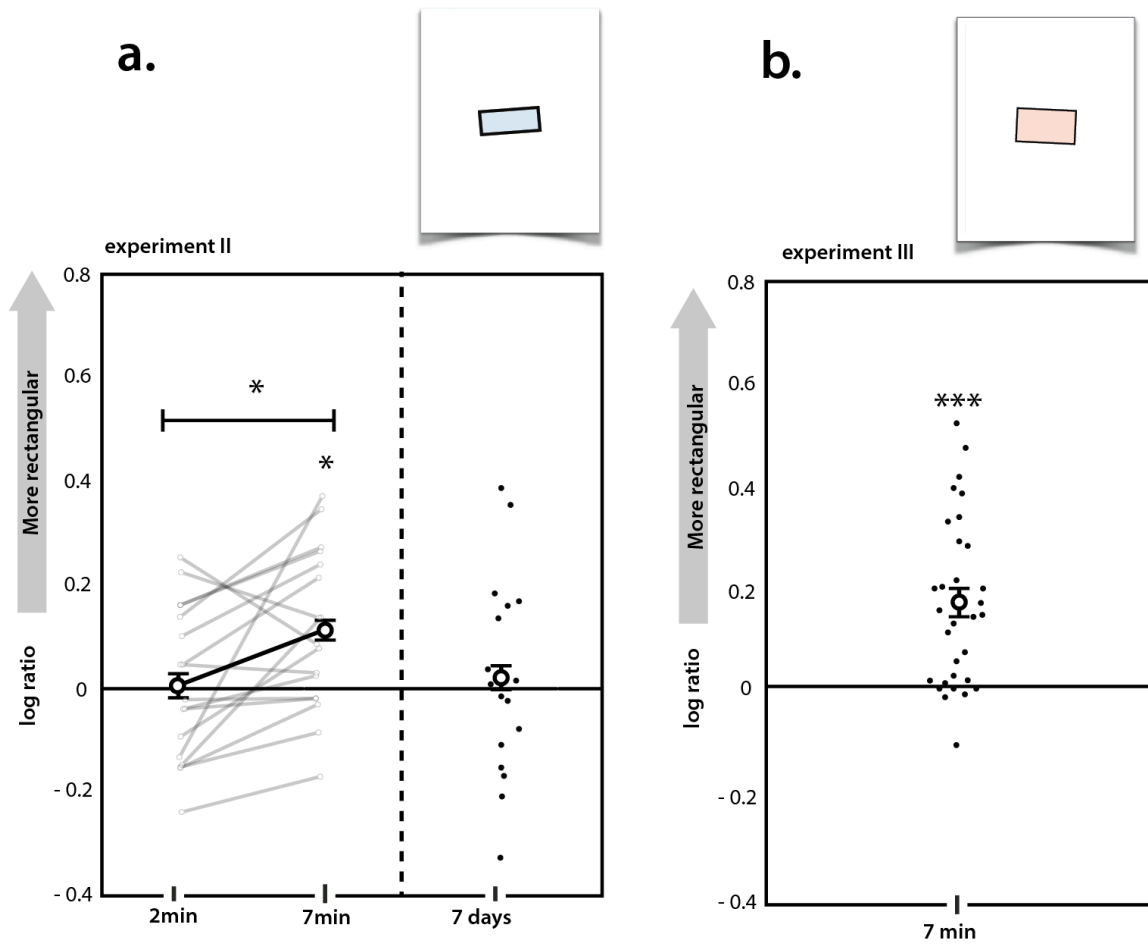
A one-sample two-sided t-test was performed to demonstrate statistical differences.



**Figure 3.2: Timeline in Experiment 2.** Each subject first drew a rectangle (prototype). The memoranda were then individually constructed to be 20% more rectangular than each participant’s prototypical rectangle. These memoranda were shown to participants, who then had to draw what they remembered at 2 minutes, 7 minutes and 7 days after the exposure.

## Results

Seventeen participants completed Experiment 2. The effects were still present in the absence of a square, but the effect arose only after a delay that was sufficient to prevent nearly perfect memory. In particular, we found 8/17, 12/17 and 9/17 participants drew an exaggerated rectangle at 2 minutes, 7 minutes and 7 days respectively. There was initially no significant exaggeration of the ratio at 2 minutes delay ( $t(16) = 0.15$ ,  $p > 0.05$ ) but then became significantly more exaggerated at 7 minutes delays ( $t(16) = 2.93$ ,  $p = 0.01$ ,  $d = 0.71$ ). We did not find an exaggeration of memory at 7 days delay ( $t(16) = 0.47$ ,  $p > 0.05$ ). However, planned t-tests showed a significant increment in the exaggeration effect between 2-min versus 7-min delay ( $t(16) = 2.83$ ,  $p = 0.012$ ,  $d = 0.69$ ), and this difference survived a Bonferroni correction.



**Figure 3.3: Results from Experiment 2 and 3.** a) Results from Experiment 2 in which we removed the square to demonstrate that the effect was not solely driven by the presence of the square. There were significant exaggerations of aspect ratio after 7-minutes, though not when tested immediately (at 2 minutes). The exaggeration at 7-minutes was also significantly larger than at the 2-minute delay. b) Results from Experiment 2I. The participants were shown and had to remember a 20% less rectangular shape than the ones they drew initially (their prototype). Participants continued to exaggerate the aspect ratio (rectangularity).

The fact that the effect grew slightly larger with longer delays within the first 10 minutes implies that the process of distinctiveness exaggeration continues after initial memory acquisition. The fact that the amplification effect was not present after 7 days is not surprising because we predicted that as the memory became very imprecise, participants would likely revert to their prototypical rectangle (i.e., for those subjects who no longer remember the specific rectangle that we gave, they will draw something close to their prototypical rectangle). Thus, the amplification effect seems to grow stronger as memory became less precise but then disappears when the memory is too weak.

### **Experiment 3**

The results from first two experiments demonstrated that the reproduced rectangles were more extreme than the rectangle that they had to remember, even if that initial rectangle was already 1.2 times as rectangular as their prototype rectangle. One possibility is that participants compared the to-be-remembered rectangle to their own prototypical rectangle (i.e., “skinner” than a usual rectangle). In this case, we would expect that the distortion would flip if participants are asked to remember a rectangle that is more “square” than their prototypical rectangle. However, another possibility is that the difference between the length of long and short sides was exaggerated. If this is the case, the rectangle would continue to be remembered as more rectangular even if it was more square than their prototype, since the prototype rectangle would not play a significant role in the encoding of the memory item.

### **Methods**

#### *Participants*

We predicted that the effect size will be similar to the result from experiment 2 (i.e. serve as internal replication experiment, a minimum number of participants of 32,  $d = 0.71$ ,  $\alpha = 0.05$ , power = 0.8). Thirty-two healthy volunteers (18 female, mean age of  $20.9 \pm 0.3$  S.E.M.) from the University of California San Diego (UCSD) community participated in the experiment. All procedures were approved by the UCSD Institutional Research Board. All participants reported normal or corrected-to-normal vision.

#### *Stimuli*

We repeated the stimulus preparation step from Experiment 1 and 2 (acquired individual’s prototypical rectangles a few weeks before the memory experiment) but created a 20% *less* exaggerated rectangles for each participant (0.8x) as opposed to the more exaggerated rectangle (1.2x) of the previous experiments.

#### *Procedure*

The same instructions were given to the participants. Then, participants had to perform the distraction task (passive listening to a story). Memory was tested 7 minutes after memory encoding.

### *Analysis*

A one-sample two-sided t-test was performed to demonstrate statistical differences.

## **Results**

Participants still drew more rectangular figures than the ones they were initially shown (26 out of 32 participants,  $t(31) = 5.74$ ,  $p < 0.001$ ,  $d = 1.01$ ; Figure 3B) even when the memoranda were less rectangular than their own prototypical rectangles. This implies that the participants did not use their prototypical rectangle as the reference (i.e. neither implicitly or explicitly; e.g., they likely did not say to themselves “I remembered that it was a particularly skinny rectangle”). Thus, the most likely explanation is that the rectangles were encoded by considering the lengths of the longer and shorter sides separately, and the difference between the longer and shorter sides of the shape were contrasted and subsequently amplified.

## **Experiment 4**

It is possible that the aspect ratio of a rectangle is a unique feature and this same principle does not apply more broadly. Thus in Experiment 4, we ask whether this phenomenon of exaggerating the distinction between features is a more general principle of memories by testing memory for relative size rather than shape. Participants needed to remember the size of two squares, and then draw them after a delay. We then asked if the distinction between them-- the size difference --was systematically exaggerated by participants in a drawing-based recall task.

## **Methods**

### *Participants*

The memory of the size might be more volatile than the ratio of the shape (e.g., the size of the objects in the real world can vary a lot depend on the viewing distance while the shape is relatively more stable). On contrary, having two contrasting objects might yield higher effect size as we found in experiment 1. We expected the effect size to be between experiment 1 vs. experiment 2 & 3 ( $d$  at 1.2 with a minimum number of participants of 12). Twelve healthy volunteers (8 female, mean age of  $21.1 \pm 0.4$  S.E.M.) from the University of California San Diego (UCSD) community participated in the experiment. All procedures were approved by the UCSD Institutional Research Board. All participants reported normal or corrected-to-normal vision.

### *Stimuli*

Our memorandum consisted of a 1.4x1.4 inch and a 1.8x1.8 inch square (one in blue and one in red color, randomly assigned) on a standard 8.5x11 inch white piece of paper with black borders. The squares were placed in the upper left and lower right corners to avoid direct side-by-side comparison.

### *Procedure*

The same instructions were given to the participants, followed by memory encoding and then a distraction task. Memory was tested with a 1-minute delay.

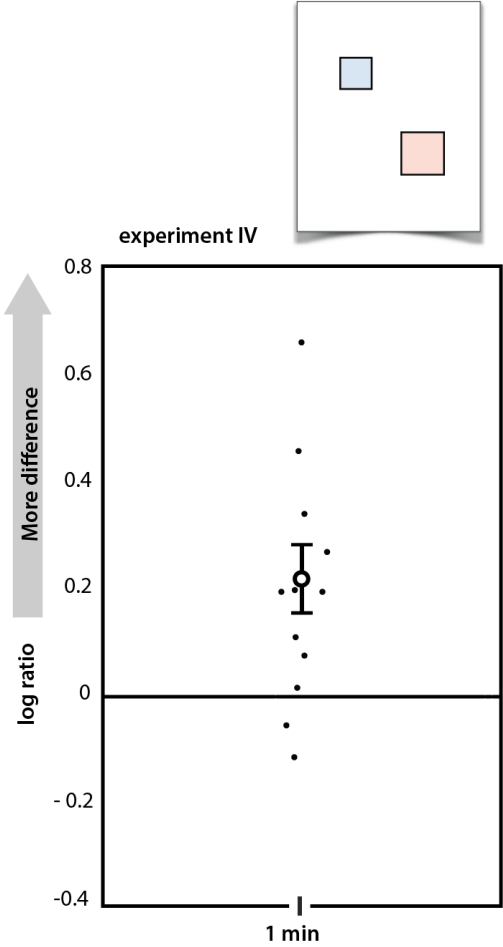
### *Analysis*

We computed the ratio between the surface area of the big and small square. We did this both for the original pair ( $1.8 \times 1.8 / 1.4 \times 1.4 = 1.653$ ) and for the ones drawn by each participant. The degree of size-difference amplification was quantified as the log of the ratio between the ratio of the original pair vs. the one the participants drew. A one-sample two-sided t-test was performed to test statistical differences.

## **Results**

When recalling the size of two squares, participants exaggerated the size difference between them by about 24% percent (10 out of 12 participants,  $t(11) = 3.019$ ,  $p = 0.012$ ,  $d = 0.87$  Figure 4). This shows

that the exaggeration of the ratio of rectangles is not unique. In another situation where participants needed to encode the relationship between two items (the size of the squares as opposed to the length of the sides of the rectangle), this difference was once again exaggerated, suggesting that distinctiveness exaggeration is a more general phenomena, and that representations are in fact distorted in memory to amplify distinctive features.



**Figure 3.4: Results from Experiment 4.** The subjects were shown two squares of different sizes. They showed an exaggeration of the size difference after a 1-minute delay.

## **Discussion**

Many previous studies have shown that visual memory tends to be systematically distorted toward the gist or prototype (e.g., Huttenlocher et al., 1991, 2000). However, if one of the goals of our memory system is to avoid confusion, then when we need to remember one item or to distinguish two similar items, memories may instead become more distinct from other memories or from the category rather than more similar. The current experiments demonstrate that, at least in some circumstances, our visual memories are in fact systematically distorted such that the unique (diagnostic) features of an item are amplified. Additionally, this amplification grows stronger with longer delays, suggesting that the process of distinctiveness exaggeration continues after initial memory acquisition. Importantly, these results show that caricature effects (Lee et al., 2000; Mauro & Kubovy, 1992) are not solely due to the way stimuli are matched to otherwise veridical memories. Instead, memories themselves are sometimes distorted to exaggerate their distinctiveness.

### **Why exaggeration?**

All memory must be a compromise between the opposing tendencies of attraction to the category center, to make the best use of gist information, and repulsion from the category center (exaggeration), to keep items distinct from each other. Neither overgeneralization nor total focus on learning a specific exemplar can result in optimal learning (Xu & Südhof, 2013). The use of these different strategies for encoding memories therefore likely differs by circumstance. In some circumstances, where we wish to recognize one object in particular as distinct from other similar objects (i.e., specificity is the goal), blending that particular object with similar ones will not allow us to achieve our goal, even if it minimizes our average error for remembering many items (e.g., Huttenlocher et al. 2001). Instead, to maintain maximum specificity for a particular item, memory should be biased in the opposite direction - amplifying the diagnostic features of an object or repulsing it from the category or gist. This should result in a caricature version of that object instead of a version that is blended together with the other similar objects.



Here we show this repulsion bias occurs in visual memory by using a recall task, thus ensuring that it is not the recognition matching process that benefits from such caricature but actually the underlying representation that is affected. In addition, we show that this occurs in memory, not perception since it increases over the course of the first several minutes. We also show this is a form of distinctiveness exaggeration rather than purely a relational encoding effect (e.g., repulsion bias) because it occurs even when only one stimulus must be remembered. However, this effect is closely related to previous work which shows repulsion biases when two closely related items are perceived. For example, when a participants look at two intersecting lines with similar orientations or two similar motion directions (either simultaneously or sequentially), they often perceive the angle between them as wider than it is (e.g., O'Toole & Wenderoth, 1977; Schwartz, Hsu, & Dayan, 2007; Marshak & Sekuler, 1979; Rauber & Treue, 1998) as well as other visual features e.g., spatial frequency (Klein, Stromeyer, & Ganz, 1974), curve line (Gibson, 1933) and oval shapes (Sweeny, Grabowecky, & Suzuki, 2011). Recent studies demonstrated a close relationship between the perceptual repulsion bias and perceptual discriminability (Wei & Stocker, 2015, 2017). This is a perceptual analog of the distinctiveness exaggeration effect we report here, and of the repulsion that results from this distinctiveness exaggeration when encoding two stimuli in memory (e.g., Experiment 1 & IV). Together, our data suggest that while attraction toward the center of the category may be the most common result -- when people need to remember many items, this helps promotes generalization of information to inform many items at once -- exaggeration or repulsion away from the center of the category may promote specificity in circumstances where one or few items need to be remembered as distinct from a category or set of items.

## **Conclusion**

Memory is a constructive and adaptive process. Instead of maintaining information as accurately as possible, the content of our memory takes into account the relation between items, our pre-existing knowledge, and our current reference frame. By studying visual memory for simple shapes and sizes, here we demonstrate that memory is not always biased toward the prototype or center of a category. On the contrary, we show that certain diagnostic features are amplified in memory, presumably to make these

items more distinctiveness from the rest of the items in that category or from similar items that we have encountered. This trajectory of distinctiveness exaggeration even continues after initial encoding. Thus, we propose that amplification of distinctive features will occur when two or more things are distinct but potentially confusable, and we are tasked with remembering one of them in particular. This may help our brains reduce confusability between items when our goal is specificity, rather than remembering many items as accurately as possible. It is a curious thought that when you walk around acquiring new memories, hours afterward, your memories are not a replica, but are instead a caricatured version of the world.

### **Acknowledgement**

Chapter 3, in full, is a reprint of the material as it appears in PsyArxiv, 2018, Chunharas, Chaipat; Brady, Timothy; Ramachandran, Vilayanur. The dissertation author was the primary investigator and author of this paper.

## References

1. Alvarez, G. A. (2011). Representing multiple objects as an ensemble enhances visual cognition. *Trends in Cognitive Sciences*, *15*(3), 122–131.
2. Bartlett, F. C. (1932). Some Experiments on the Reproduction of Folk-Stories. *Folklore*, *31*(1), 30–47.
3. Brady, T. F., & Alvarez, G. A. (2011). Hierarchical encoding in visual working memory: ensemble statistics bias memory for individual items. *Psychological Science*, *22*(3), 384–392.
4. Dubé, C., Zhou, F., Kahana, M. J., & Sekuler, R. (2014). Similarity-based distortion of visual short-term memory is due to perceptual averaging. *Vision Research*, *96*, 8–16.
5. Freyd, J. J., & Johnson, J. Q. (1987). Probing the time course of representational momentum. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *13*(2), 259–268.
6. Gibson, J. J. (1933). Adaptation, after-effect and contrast in the perception of curved lines. *Journal of Experimental Psychology*, *16*(1), 1.
7. Hemmer, P., & Steyvers, M. (2009). A Bayesian account of reconstructive memory. *Topics in Cognitive Science*, *1*(1), 189–202.
8. Huang, J., & Sekuler, R. (2010). Distortions in recall from visual memory: two classes of attractors at work. *Journal of Vision*, *10*(2), 24.1–27.
9. Hubbard, T. L., Hutchison, J. L., & Courtney, J. R. (2010). Boundary extension: findings and theories. *Quarterly Journal of Experimental Psychology*, *63*(8), 1467–1494.
10. Huttenlocher, J., Hedges, L. V., & Duncan, S. (1991). Categories and particulars: prototype effects in estimating spatial location. *Psychological Review*, *98*(3), 352–376.
11. Huttenlocher, J., Hedges, L. V., & Vevea, J. L. (2000). Why do categories affect stimulus judgment? *Journal of Experimental Psychology. General*, *129*(2), 220–241.
12. Intraub, H., & Richardson, M. (1989). Wide-angle memories of close-up scenes. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *15*(2), 179–187.
13. Klein, S., Stromeyer, C. F., 3rd, & Ganz, L. (1974). The simultaneous spatial frequency shift: a dissociation between the detection and perception of gratings. *Vision Research*, *14*(12), 1421–1432.
14. Lee, K., Byatt, G., & Rhodes, G. (2000). Caricature effects, distinctiveness, and identification: testing the face-space framework. *Psychological Science*, *11*(5), 379–385.
15. Marshak, W., & Sekuler, R. (1979). Mutual repulsion between moving visual targets. *Science*, *205*(4413), 1399–1401.
16. Mauro, R., & Kubovy, M. (1992). Caricature and face recognition. *Memory & Cognition*, *20*(4), 433–440.

17. O'Toole, B., & Wenderoth, P. (1977). The tilt illusion: repulsion and attraction effects in the oblique meridian. *Vision Research*, *17*(3), 367–374.
18. Rauber, H. J., & Treue, S. (1998). Reference repulsion when judging the direction of visual motion. *Perception*, *27*(4), 393–402.
19. Schacter, D. L., Guerin, S. A., & St. Jacques, P. L. (2011). Memory distortion: an adaptive perspective. *Trends in Cognitive Sciences*, *15*(10), 467–474.
20. Schwartz, O., Hsu, A., & Dayan, P. (2007). Space and time in visual context. *Nature Reviews. Neuroscience*, *8*, 522.
21. Spencer, J. P., & Hund, A. M. (2002). Prototypes and particulars: geometric and experience-dependent spatial categories. *Journal of Experimental Psychology. General*, *131*(1), 16–37.
22. Sweeny, T. D., Grabowecky, M., & Suzuki, S. (2011). Simultaneous shape repulsion and global assimilation in the perception of aspect ratio. *Journal of Vision*, *11*(1), 16.
23. Tanaka, J. W. (1996). Caricature Recognition in a Neural Network. *Visual Cognition*, *3*(4), 305–324.
24. Wei, X. X., & Stocker, A. A. (2015). A Bayesian observer model constrained by efficient coding can explain 'anti-Bayesian' percepts. *Nature neuroscience*, *18*(10), 1509.
25. Wei, X. X., & Stocker, A. A. (2017). Lawful relation between perceptual bias and discriminability. *Proceedings of the National Academy of Sciences*, *114*(38), 10244-10249.
26. Xu, W., & Südhof, T. C. (2013). A neural circuit for memory specificity and generalization. *Science*, *339*(6125), 1290–1295.
27. Yassa, M. A., & Stark, C. E. L. (2011). Pattern separation in the hippocampus. *Trends in Neurosciences*, *34*(10), 515–525.