From the holdings of the Department of Special Collections , Princeton University Library, Princeton, NJ

These documents are either in the public domain, or can only be used for educational and research purposes (Fair use) as per U.S. Copyright law (text below). By accessing this file, all users agree that their use falls in the public domain or within fair use as defined by the copyrightlaw. We do not charge any permission or usage fees for the publication of images of material inour collections, including those provided by Princeton University Library, via our website, catalog, or directly from staff.For more information regarding copyright, credit, and citation please review the information onthe Copyright, Credit and Citation Guidelines page of our website.
https://library.princeton.edu/special-collections/policies/copyright-credit-and-citation-guidelines. If you have any questions, including how to order higher-quality images from the collection, contact Special Collections Public Services via our Ask Us! Form. https://library.princeton.edu/special-collections/ask-us

---

Doctoral Dissertations, 2011-2026 Psychology

2016

Role of cognitive actions in learning

Fan, Judith

Princeton, NJ : Princeton University

---

http://arks.princeton.edu/ark:/88435/dsp01pc289m48w

U.S. Copyright law: The copyright law of the United States (Title 17, §108, United States Code) governs the making of photocopies or other reproductions of copyrighted material. Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specified conditions is that the photocopy or reproduction is not to be used for any purpose other than private study, scholarship, or research. If a user makes a request for, or later uses, a photocopy or reproduction for purposes in excess of fair use that user may be liable for copyright infringement. This institution reserves the right to refuse to accept a copying order if, in its judgment, fulfillment of the order would involve a violation of copyright law. https://www.copyright.gov/title17/title17.pdf

Role of cognitive actions in learning

Judith E. Fan

a dissertation
presented to the faculty
of princeton university
in candidacy for the degree
of Doctor of Philosophy

recommended for acceptance
by the Department of
Psychology

Advisor: Nicholas B. Turk-Browne

May 2016

© Copyright by Judith E. Fan, 2016. All rights reserved.

Abstract

Just as the flight of the honeybee links flower to flower, multiplying their numbers across generations, acts of cognition connect disparate ideas and engender new ones. We promote these connections in various ways, by selecting which thoughts to bring to mind, and actively exploring the world around us. My research program seeks to uncover how we employ such cognitive actions to make sense of our world.

This notion of *cognitive actions* refers broadly to operations that have the capacity to transform mental representations through their very execution. In some cases, these acts are not readily visible by external observers (e.g., memory retrieval). In other cases, these acts are accompanied by motor movements that are externally observable (e.g., reaching) or leave a visible trace (e.g., drawing). Understanding the consequences of such cognitive actions is especially important because they are the essence of intelligent, goal-directed behavior. My dissertation focuses on object learning — that is, how people acquire knowledge about the diagnostic features of objects and adaptively exploit this knowledge to achieve their goals.

In Chapters 1-3 I will describe a series of studies I conducted to examine the consequences of internally-directed cognitive actions (i.e., internal attention) on the organization of long-term memory representations, as well the influence of these memories on attention and perception. Chapter 4 will address the influence of task-related and sensory feedback (i.e., visual and haptic) during task practice on perceptual representations. Finally, in Chapter 5, I describe investigations of the impact of 'visual production' (i.e., drawing) on the mental representation of object categories.

Contents

Abstract 1

0 Introduction 1

1 How does attending to an object in mind guide visual learning? 4
- 1.1 Experiment 1: Retrieval History and the Tuning of Object Representations . 12
- 1.2 Experiment 2: Ruling Out Alternative Explanations 22
- 1.3 Experiment 3: Variable vs. Focused Retrieval History 27
- 1.4 Modeling Accessibility and Precision 32
- 1.5 General Discussion 37
- 1.6 Conclusions 44

2 How do visual memories affect what we pay attention to? 46
- 2.1 Experiment 1: Guidance in Visual Search 48
- 2.2 Experiment 2: Attentional Capture from VLTM 55
- 2.3 General Discussion 62

3 How do visual memories affect our perceptual judgments? 65
- 3.1 Methods 68
- 3.2 Results 75
- 3.3 Discussion 86

4 What kinds of feedback enhance perceptual learning? 89
- 4.1 Experiment 1: Centroid Discrimination 93
- 4.2 Experiment 2: Center-of-Mass Discrimination 107
- 4.3 Experiment 3: Role of Pointing Movements in Learning 114
- 4.4 Experiment 4: Change in the Fidelity of Representation 118
- 4.5 General Discussion 121

5 Visual production as a window into perception and learning 130
5.1 How do visual recognition and production support communication? . . . 132
5.2 Experiment 1: Recognizing pictures of objects 132
5.3 Experiment 2: Producing drawings of objects 138
5.4 Experiment 3: Does visual exposure to sketches suffice to induce differentiation? 153
5.5 Experiment 4: Does observing sketch production induce learning? 155
5.6 Discussion 156

6 Conclusion 163
6.1 From cognitive actions to cognitive technologies 166

References 188

The following authors contributed to Chapters 1 and 2: Judith E. Fan and Nicholas B. Turk-Browne. Chapter 1 was published in 2013 in the journal *Cognition*, under the title: "Internal attention to features in visual short-term memory guides object learning." Chapter 2 was published in 2015 in the journal *Journal of Experimental Psychology: Learning, Memory, & Cognition* under the title: "Incidental biasing of attention from visual long-term memory."
The following authors contributed to Chapter 3: Judith E. Fan, J. Benjamin Hutchinson, and Nicholas B. Turk-Browne. Chapter 3 is currently *in press* at *Journal of Vision*, under the title: "When past is present: Substitutions of long-term memory for sensory evidence in perceptual judgments."
The following authors contributed to Chapter 4: Judith E. Fan, Nicholas B. Turk-Browne, and Jordan A. Taylor. Chapter 4 was published in 2015 in the journal *Journal of Experimental Psychology: Human Perception & Performance*, under the title: "Error-driven learning in statistical summary perception."
The following authors contributed to Chapter 5: Judith E. Fan, Daniel L. K. Yamins, and Nicholas B. Turk-Browne. A subset of this work was presented at the 37th Annual Meeting of the Cognitive Sciences Society in 2015, and reprinted as part of the conference proceedings under the title: "Common object representations for visual recognition and production." Some passages were excerpted from a 2015 publication in the journal *Translational Issues in Psychological Science*, under the title: "Drawing to learn: how producing graphical representations enhances scientific thinking."

This work was supported by NSF GRFP DGE-0646086, NIH R01 EY021755, a Program in Cognitive Science Graduate Fellowship, and the David A. Gardner '69 Magic Project at Princeton University.

Listing of figures

1.1 Task display. On each trial of the practice phase, observers saw an object defined by a unique color, orientation, and location. After offset, observers were prompted by a post-cue ("xxx" replaced by "color", "orientation", "location", or "click") to report one of its features from VSTM. Observers reported this feature by continuously adjusting a memory probe until it matched the original stimulus. All objects were viewed multiple times during the practice phase. This was followed by a test phase in which observers could rely only on VLTM to report the features of a probe. The nature of the probe differed across experiments. (A) In Experiment 1, the memory probe was a black, canonically oriented version of an object presented in the center of the display. Observers used a visible response wheel to manipulate the memory probe along the cued dimension. (B) In Experiments 2 and 3, the memory probe was identical to the original object in appearance, except for the cued dimension. Observers adjusted the probe by moving the mouse until a good match was found. There was no response wheel, and the mapping between mouse position and feature space was perturbed from trial-to-trial. 11

1.2 (A) Design. In Experiments 1 and 2, all objects were viewed three times during the practice phase. Each object was assigned to the Constant, Switch, or Passive condition. For Constant objects, the same feature was reported throughout both the practice and test phases. Switch objects were probed on the same feature in the practice phase but on an unpracticed feature in the test phase. Passive objects were viewed the same number of times during the practice phase, but no features were reported; one of the unpracticed features was subsequently tested. (B) In Experiment 3, all objects were viewed twice during the practice phase. Each object was assigned to the Constant, Switch, or Variable condition. Constant and Switch conditions were identical to before (but with two repetitions). For Variable objects, two different features were retrieved in the practice phase, and like the Switch condition, an unpracticed feature was subsequently tested. . 14

1.3 Results of Experiment 1. (A) Empirical distributions depicting the probability of observing an error of a given magnitude in the VLTM test for each condition. These distributions were constructed for visualization purposes only, by pooling data across all observers. (B) Task performance across VSTM practice and VLTM test phases as measured by RMSE for objects successfully reported during practice phase. In the test phase, practiced features (Constant) were reported most accurately, while unpracticed features of practiced objects (Switch) were impaired relative to unpracticed features of unpracticed objects (Passive). The height of the dotted line represents expected performance under random guessing. Error bars reflect ±1 SEM. *p<0.05, ***p<0.001. 21
1.4 Results of Experiment 2. (A) Empirical distributions for each condition in the VLTM test phase, pooling data across observers for visualization purposes. (B) Task performance across VSTM practice and VLTM test phases as measured by RMSE. In the test phase, unpracticed features of practiced objects were impaired (Switch), even when observers received equal (vs. Passive) or greater (vs. Constant) veridical sensory exposure to these features during the practice phase. Error bars reflect ±1 SEM. □□$p < 0.01$. 28
1.5 Results of Experiment 3. (A) Empirical distributions for each condition in the VLTM test phase, pooling data across observers for visualization purposes. (B) Task performance across VSTM practice and VLTM test phases as measured by RMSE. In the test phase, unpracticed features were impaired when belonging to an object for which two (Variable) vs. one (Switch) other features had been practiced. Error bars reflect ±1 SEM. **p<0.01. 31
1.6 Model Fitting Results. (A) Guessing (g) and precision (1/sd) across conditions in the VLTM test phase of each experiment. The probability of guessing was generally reduced for practiced (Constant) vs. unpracticed feature (Switch/Variable/Passive) features, but the precision of non-guesses did not significantly differ between conditions. Error bars reflect ±1 SEM. * $p < 0.05$, † $p < 0.10$. (B) Relationship between parameter estimates and primary RMSE measure in Experiment 1. There were robust positive correlations between RMSE and g, but unreliable correlation between RMSE and sd. This pattern of correlations held across all experiments. 36

1.7 Interpretation. (A) Initially, attentional weights allocated to each feature dimension when forming a representation are roughly balanced. (B) Cued retrieval to a specific feature dimension may disrupt this equal weighting by inducing competition between dimensions for representation in short-term memory. This competition may be resolved via the selection of the object's value on the cued dimension and inhibition of values from uncued dimensions. Crucially, the consequences of this selective internal attention are not restricted to short-term memory, but extend to how various feature dimensions of an object are prioritized in long-term memory. (C) If the same feature is retrieved again from short-term memory in a subsequent encounter with that object, the competition among dimensions may be less fierce because this feature is already more strongly represented. As a result, the representation of the uncued dimensions may be further suppressed only slightly. (D) If a different feature is cued upon a subsequent encounter, accurate retrieval of this cued dimension from short-term memory may be more difficult due to the stronger status of the previously retrieved feature and prior inhibition of the newly cued feature. Moreover, given the weaker status of the cued feature, other previously uncued features may compete more, resulting in the further marginalization of still uncued feature dimensions in long-term memory. 45

2.1 On each trial, one shape was presented in a unique color and location. After offset, a post-cue ("xxx" replaced by "color" or "location") prompted observers to report one of these two features by continuously adjusting a memory probe until it matched the original stimulus. Objects were viewed multiple times and the same feature was post-cued each time. 47

2.2 Search phase of Experiment 1. Observers determined whether a cued shape was present or absent within a cluttered display, regardless of its color or location (bottom). Mean log-transformed search RTs for correctly detected targets, as a function of whether the target matched the color and/or location of the cued shape from encoding (top). Error bars reflect ±1 S.E.M. 54

2.3 Dual-task phase of Experiment 2. (A) While maintaining a (monochromatic) shape in working memory, observers discriminated the orientation (clockwise/counterclockwise) of the tilted Gabor (target). The color of this patch was associated with the shape on half of trials (valid). On the remaining trials, the vertical Gabor (distractor) appeared in this color (invalid). (B) Mean orientation-discrimination RTs as a function of validity. Error bars reflect ±1 S.E.M. 60

3.1 Task Display. (a) During initial exposure trials, a colored shape was briefly presented and then an achromatic version of the shape reappeared to probe observers for the color. (b) Final test phase trials were identical, other than the use of forward and backward masks, the duration of the initial shape, and the fact that the shape had previously appeared in a different color. 70
3.2 Modeling Approach. Responses across observers in the final test phase were modeled as a mixture of three component distributions: a uniform distribution to capture guess responses and von Mises distributions to capture responses centered on the 'original' color of the shape from exposure and its 'current' color, respectively. β = probability of anchoring on a color; σ-1 = precision of anchored responses. 73
3.3 Distribution of Errors in Test Phase. Each panel contains a histogram of errors pooled across all observers and the corresponding fit from the three-component model (dotted line). In (a) and (b), all visible trials are plotted: in (a), errors represent the difference between each response and the current color, whereas in (b) errors represent the difference between each response and the original color. (c) On invisible trials, responses clustered around the original color. 76
3.4 Modeling Results for Test Phase. (a) The probability of anchoring on the current color and the precision of these responses decreased as stimulus duration was reduced from 150ms to 50ms to 10ms. (b) This reduction increased the probability of anchoring on the original color. At shorter durations (50ms, 10ms), the precision of responses anchored on the original color was high, at least relative to the precision for the original color on initial exposure trials (see Table 1). (c) On trials containing no colored shape, anchoring on the original color occurred reliably, and at a consistent rate across experiments, with similar and high precision. Error bars represent bootstrapped 68% confidence intervals (comparable in width to 1 s.e.m). 80
3.5 Distribution of responses relative to the current and original colors on visible test trials. Errors were normalized by taking the difference between the response and current color and dividing it by the distance between the original color and current color for each item. This results in a distribution in which the current and original colors have been remapped to 0 and 1, respectively, and errors have been scaled in proportion. 83

3.6 Model Comparison Results. (a) On visible test trials, the three-component mixture model ("swap") provided the best overall fit to the data across experiments, relative to reduced models containing only the original-color component or only the current-color component, plus the uniform component. Differences in AIC are plotted relative to the null model containing only the uniform component. The blend model, while performing well, did not exceed the swap model in any experiment. b) On invisible test trials, where there was no current color, the two-component model containing the original-color and uniform components significantly outperformed the null model. The performance of the swap model was comparable, except for the penalty due to the additional parameter. Error bars indicate 68% confidence intervals. 85

4.1 Figure 1: Distribution of centroid locations in the quadrant task followed a hyperbolic function: $r = \frac{c}{\sqrt{\cos(\theta)\sin(\theta)}}$, where r is the radial distance from the center of the display, θ represents randomly sampled angles, and c is a scaling factor that controls the overall distance to the nearest quadrant boundaries. Smaller values of c produce displays with centroids falling closer to boundaries, thus making discrimination more difficult. 96

4.2 Task and procedure. (A) Eight dots of uniform size presented for 200ms in random locations on each trial. During training, participants pointed to the centroid on touchscreen. During the pre-test and post-test, participants discriminated which quadrant contained the centroid. (B) Groups received different trial-by-trial feedback during training: true centroid location shown after response (Vector), point score reflecting distance to centroid (Scalar), no feedback (Control), or centroid location shown prior to responding (Guided). (C) In Experiment 2, dots appeared in random sizes and locations. (D) During training in Experiment 3, participants typed in the code of the grid cell containing the centroid, and then were presented with the location of the centroid. (G) Participants in Experiment 4 performed the pointing task during training while receiving vector error feedback; the grid task was used to obtain a more precise measure of changes in perceptual discrimination. 102

4.3 Experiment 1 Results. A) Overall performance (RMSE) on the pointing task during the training phase as a function of trial number for the Vector (Salmon), Scalar (Blue), and Control (Gray) groups. All groups performed better than chance (dashed line). B) Performance in the pointing task broken down into accuracy (radial error) and precision (standard deviation). C) Percent change in accuracy between post-test and pre-test in the quadrant task. Error bars represent 1 SEM *<0.05, <0.1. 105

4.4 Experiment 2 Results. A) Overall performance (RMSE) on the pointing task as a function of trial number for the Vector (Salmon), Scalar (Blue), and Guided (Purple) groups. All groups performed better than chance (dashed line), with Guided participants making substantially smaller errors. B) Performance in the pointing task broken down into accuracy (radial error) and precision (standard deviation). C) Percent change in accuracy between post-test and pre-test in the quadrant task. Error bars represent 1 SEM *<0.05, <0.1. 112
4.5 Experiment 3 Results. A) Overall performance (RMSE) greatly exceeded chance, where continuous error computed as distance between center of selected grid cell and the centroid. B) Performance in the grid task broken down into accuracy (radial error) and precision (standard deviation). C) Percent change in accuracy between post-test and pre-test in the quadrant task. Error bars represent 1 SEM *<0.05, <0.1. 117
4.6 Experiment 4 Results. (A) Overall performance (RMSE) greatly exceeded chance on the pointing task with Vector feedback. B) Performance in the pointing task broken down into accuracy (radial error) and precision (standard deviation). (C) Error distribution in the pre-test and post-test modeled as bivariate Gaussian distribution fit to participant errors. (D) Percent change in grid task performance (RMSE) between post-test and pre-test; negative value reflects enhanced fidelity. Error bars represent 1 SEM *<0.05, <0.1. 120

5.1 Visual recognition entails mapping an image onto a specific idea (e.g., object identity based on a photograph). Visual production entails expressing a specific idea in an image (e.g., identifiable sketch based on an object concept). (Art credit: Jeffery Thompson) 133
5.2 Multi-domain imageset containing sketches, photographs, and 3D-rendered images of 105 real-world objects. 134
5.3 (a) Feature extraction using a neurally predictive, deep convolutional neural network model optimized for performance on challenging object recognition tasks. (b) Correlation matrices for each image domain, displaying the overall layout of objects in high-dimensional feature space. Each entry shows correlation distance (1-ρ) between feature vectors for a pair of objects. 136
5.4 Cross-domain similarity between image domains increases as a function of model layer. 137

5.5 Task: On each trial, participants were prompted with a trial-unique photo (N=311) or word (N=282) that referred to a target object for them to sketch. The computer guessed the identity of the drawn object in real time, providing participants immediate feedback about the quality of their sketch. The rank of the target in the list of all 64 guesses (ordered by confidence) returned by the computer was used to track changes in performance across trials. 140
5.6 (a) Stimuli: Objects belonged to eight visual categories, each containing eight items. (b) Design: Each participant was randomly assigned two of these categories. During training, participants drew four randomly selected objects in one category (Trained) multiple times. Before and after training, participants drew the other four objects in that category (Near), as well as the objects in the second category (Far), once each. 143
5.7 Sample sketches from the experiment, with target label and model's top guess. 144
5.8 An independent cohort of human participants guessed the identity of objects depicted in drawings from the image-cue experiment. Human and computer recognition performance (d') was highly consistent across objects (r=0.649). 145
5.9 Changes in performance between post-test and pre-test for each condition: (*left*) when computing target rank among all 63 distractors and (*right*) when computing target rank relative to remaining 7 objects within the target category only. Error bars represent 1 s.e.m. 148
5.10 Feature distance (1-correlation) between successive drawings of the same object (Trained condition only). 151
5.11 Representational similarity between conditions across phases (averaged over object identity): low values in the large off-diagonal blocks correspond to larger average correlation distances between objects in different categories. 152
5.12 Representational similarity between conditions across phases (averaged over object identity): low values in the large off-diagonal blocks correspond to larger average correlation distances between objects in different categories. 155

6.1 A: Mendeleev's Periodic Table as of 1871. B: Drawings of four Galápagos finch specimens highlighting differences in beak morphology. C: A Feynman diagram depicting the exchange of one quantum between two electrons. 167
6.2 A: 'Bull – plate 11' by Pablo Picasso. B: Algorithmically-generated assembly instructions for a TV stand. From Agrawala, M., et al., 2003, ACM Transactions on Graphics, 22(3), 828. 169
6.3 Early architectural sketch of "Fallingwater" by Frank Lloyd Wright (left) with recent photograph (right). 171

To each who sees things not just as they are, but what they could be.
To my mother and father, who always have in me.

Acknowledgments

There are so many people to thank.

First, my advisor: Nick Turk-Browne. To the extent I have grown as a scientist at Princeton, I have Nick to thank, top to bottom. Thank you for modeling the discipline and rigor that exemplifies programmatic research, and the open-mindedness to nurture even my most tentative ideas toward their fruition. For helping me to think clearly and write simply, for the benefit of everyone. (May we all learn to emulate your *mellow precision*.) For taking this journey with me into the unknown — week after week, year after year, for the past fifty-eight months.

Thank you to Betsy Paluck, who took me in during my first year and helped me gain a taste for doing experimental psychology in the field, which has never left me.

Thank you to Jordan Taylor, who introduced me to the world of motor and skill learning. It has been a pleasure to learn from you, and together explore how vision and action work together.

Thank you to my other Princeton committee members: Uri Hasson and Ken Norman.

My thinking and approach to science has been profoundly shaped through my friendship and close collaboration with Dan Yamins. Your investment of knowledge, patience, and confidence has done more to embolden me to tackle the problems that matter most than anything else. To the discoveries we still have yet to make together.

Thank you to my amazing collaborators, especially Ben Hutchinson, Simon Cullen, Cameron Ellis, Julian De Freitas, and Jordan Suchow.

Thank you to labmates and other colleagues, for discussions and support: Jiaying Zhao, José Ribas-Fernandes, Natalia Córdova, Aaron Kurosu, Jon Berliner, Alexa Tompary, Ghootae Kim, Megan deBettencourt, Mariam Aly, Sam McDougle, James Antony, Justin Jungé, Michael Cohen, Arash Afraz, Maryam Vaziri Pashkam, Sukeun Jeong, Daryl Fougnie, Tim Brady, Talia Konkle, and Jason Haberman.

Thank you to my wonderful research assistants: Max Luo, Annie Chen, Jenn Rawding, Ryan O'Connell, Jordan Gunn, Rachel Klebanov, Laura Herman, and Jessica Ji.

To my dear friends Lisa Jing, Daniel Bear, and Albert Shieh, who have seen it all.

Before I came to Princeton, the Harvard Vision Lab was my home. I might have left experimental science altogether if not for the extraordinary mentorship of George Alvarez, who provided me with just the right balance of autonomy and guidance in my senior year to let me experience the thrill and heartache of scientific discovery for the very first time. I've been addicted ever since.

Thank you to Ken Nakayama, sprawling conversations with whom can only enliven and expand one's mind. The breadth of your curiosity and your incisive observations are a continual inspiration to us all.

Thank you to Barbara Tversky, for illustrating the beauty and power of gestures, sketches, maps, and diagrams for the world. From your writings, I discovered what I hope to dedicate myself to for years to come: how we make pictures with our hands to think, communicate, and create. From speaking with you, I've learned even more.

Thank you to my former teachers, especially Karen Clare, Rick Fitzgerald, Geoff Brown, Richard Peterson, Chris Marsh, and Beatrice Hecht.

And above all, thank you to my mother and father, Shu and Yuan, and my brother, Richard. I would not have made it here without you.

0

Introduction

Just as the flight of the honeybee links flower to flower, multiplying their numbers across generations, acts of cognition connect disparate ideas and engender new ones. We promote these connections in various ways, by selecting which thoughts to bring to mind, and actively exploring the world around us. My research program seeks to uncover

how we employ such cognitive actions to make sense of our world.

This notion of *cognitive actions* refers broadly to operations that have the capacity to transform mental representations through their very execution. In some cases, these acts are not readily visible by external observers (e.g., memory retrieval). In other cases, these acts are accompanied by motor movements that are externally observable (e.g., reaching) or leave a visible trace (e.g., drawing). Understanding the consequences of such cognitive actions is especially important because they are the essence of intelligent, goal-directed behavior. My dissertation focuses on object learning — that is, how people acquire knowledge about the diagnostic features of objects and adaptively exploit this knowledge to achieve their goals.

Chapters 1-3 describes three studies I conducted to examine the consequences of internally-directed cognitive actions (i.e., internal attention) on the organization of long-term memory representations, as well as the subsequent influence of these memories on visual attention and perception. In the first study, we discovered that directing internal attention to one feature of an object biased visual long-term memory (VLTM) for that object by suppressing the accessibility of its other features (e.g., if the location of a red object was practiced, its color was less likely to be remembered later). In the second study, we found that these feature associations in VLTM automatically guided attention during visual search and captured attention even when these features were no longer relevant to the task. In the third study, we examined the interaction between associations in VLTM and sensory inputs on perceptual judgments about other, similar objects.

Chapter 4 addresses the influence of task-related and sensory feedback (i.e., visual and haptic) on statistical summary perception, which refers to the ability to accurately extract statistical summary features (e.g., mean location) from groups of objects. We accomplished

this by training participants to estimate the centroid of an array of objects on a touchscreen display under different feedback conditions. Before and after training, they completed a transfer test requiring perceptual discrimination of the centroid. We found that vector error feedback, which conveys error both in terms of distance and direction, was the only form of feedback that improved perceptual discrimination of the centroid on the transfer test. Moreover, this form of feedback was effective only when coupled with reaching movements toward the visual objects. These findings suggest that sensory-prediction error — signaling the mismatch between expected and actual consequences of an action — may play a previously unrecognized role in tuning perceptual representations.

Chapter 5 reports investigations of how training on *visual production* (i.e., drawing) tasks reshapes the mental representation of visual object categories. We first examined the possibility that a common internal representation of visual objects supported both recognition and production. We found that a model of ventral visual cortex only optimized to recognize objects in photographs generalizes to drawings of objects, suggesting that the capacity for visual abstraction is rooted in the functional architecture of the visual system. Next, we tested the effects of visual production practice. On each trial, participants sketched an object. The model then guessed the identity of the sketched object, providing real-time feedback. We found that repeatedly sketched objects were better recognized after training, and this reflected differentiation between them in the model representation.

1

How does attending to an object in mind guide visual learning?

It can be tempting to compare visual perception to the registration of light as accomplished by a camera, or to liken memory retrieval to reading a file from a hard drive.

However, both of these metaphors grossly underestimate the very dynamic aspects of these operations that make these such flexible and powerful faculties of the mind. For instance, our perceptual experience is only partly determined by sensory inputs: it is also powerfully shaped by our memories and expectations (Synofzik et al., 2006, von Helmholtz, 1925). Secondly, our memories are not inert: recovering a memory may actually alter it, affecting subsequent retrieval of the same information (Anderson et al., 1994, Suengas & Johnson, 1988). A core aim of my research in this area is to elucidate the mechanisms by which remembering and seeing mutually influence each other during interactions with objects we frequently encounter.

1.0.1 How does internal attention guide learning?

Physical actions bring about lasting changes to objects we encounter. These interventions may be more consequential in some cases (e.g., transforming a forest into pulp) than others (e.g., thumbing through a new book), but totally inconsequential actions are rare. Just as physical objects bear traces of such encounters, mental objects may reflect their history of manipulation by the mind. This notion of 'cognitive actions' refers broadly to processes that manipulate mental representations in various ways. Understanding the consequences of such actions is especially important because they are the essence of human cognition (Andrews-Hanna et al., 2010, Klinger & Cox, 1987, Mason et al., 2007, Singer, 1966). As a case study of cognitive actions, here we investigate the consequences of reflective, or internal attention (Chun et al., 2011b, Chun & Johnson, 2011).

Attention typically refers to the prioritization of sensory information that is either inherently salient (Theeuwes, 1992, Treisman & Gelade, 1980) or relevant to current goals (Most

et al., 2005). Deploying attention to certain information and not other information does not merely determine what gets processed downstream (Al-Aidroos et al., 2012), but can also alter the perceptual experience of this information (Carrasco et al., 2004). There are both benefits and costs of attention: processing of selected information can be facilitated (Posner et al., 1980) and processing of unselected information can be inhibited (Houghton & Tipper, 1994).

These forms of attentional modulation can have longer-term consequences for memory. On the one hand, allocating attention to an item during encoding enhances its later recognition (Chun & Turk-Browne, 2007, Uncapher et al., 2011). On the other hand, removing attention from an item during encoding not only worsens recognition, but can actually produce a memory cost for this item relative to novel items (Fox, 1995a, Tipper & Cranston, 1985). This cost may reflect an attentional weighting mechanism that actively inhibits distractors, shaping the representation of an item in memory based on current task goals (Lavie & Fox, 2000, Neill & Valdes, 1992).

The impact of attention on memory has been investigated mostly in the case of external attention, which entails the selection of representations that are directly supported by sensory information, and that do not require maintenance in working memory to remain active. Here we investigate the mnemonic consequences of internal attention, which we define as the selection of representations that are being maintained temporarily in working memory, and that are no longer supported by externally available sensory information. External attention might nevertheless provide a useful starting point for thinking about how internal attention affects memory. This analogy is supported by the fact that external and internal attention rely on similar mechanisms: (1) they are both highly selective and capacity limited

(Chun et al., 2011a), (2) deploying one form of attention interferes with the other (Awh & Jonides, 2001), and (3) both forms of attention engage overlapping networks of brain regions (Nobre et al., 2004a).

1.0.2 Consequences of Memory Retrieval

Given these commonalities to external attention, memory may be enhanced when internal attention is allocated to an item and suppressed when it is removed from an item. Support for this prediction comes from research on retrieval from long-term memory. While overly simple conceptions of remembering liken retrieval to finding and reading a file from a hard drive, the act of recovering the contents of a representation from long-term memory actually alters these contents and affects subsequent retrieval of the same information.

These alterations can be beneficial. For example, testing of recently learned information — such as a history lesson (Nungester & Duchastel, 1982) or word pairs (Carpenter et al., 2006, Carrier & Pashler, 2002) — promotes long-term retention of this information. Moreover, retention is better after repeated testing (i.e., after multiple retrieval attempts) than after an equal number of opportunities to study the same material (Karpicke & Roediger, 2008, Roediger & Karpicke, 2006a). Relatedly, information tends to be better remembered when it has been generated from internal knowledge than when acquired via external sources (Crutcher & Healy, 1989, Slamecka & Graf, 1978).

Memory retrieval can also have deleterious effects. For example, recall of one item from memory makes other related items less accessible for later recall (i.e., retrieval induced forgetting; (Anderson et al., 1994, 2000). Specifically, in such studies, participants study a list of category-exemplar word pairs (e.g., Fruit-Grape, Fruit-Peach, Mammal-Cow), practice re-

calling a subset of the exemplars with a category cue (e.g., Fruit-G—?), and finally, perform a recollection test of all exemplars. Memory is better for the pairs practiced in the second phase (e.g., Fruit-Grape) relative to unpracticed pairs (e.g., Mammal-Cow), consistent with the testing effects described above. Among unpracticed pairs, however, memory is worse for those exemplars that share a category cue with a practiced exemplar (e.g., Fruit-Peach). These results are interpreted as evidence that inhibitory mechanisms suppress memory of competitors during initial recall.

1.0.3 Internal Attention and Visual Learning

Previous work has focused on the retrieval of stable information from long-term memory, such as semantic categories and their members. However, similar competitive processes may operate in a short-term store of recently experienced visual information, when a subset of this information is selected via internal attention. This may help shape how items are represented during initial encoding, and ultimately guide longer-term learning about these items. This view is compatible with modal models of memory, which emphasize the importance of a short-term store as the nexus between ongoing perception and long-term knowledge (Shiffrin & Atkinson, 1969b). Indeed, the hallmark of short-term memory is that representations can be manipulated in the service of ongoing behavior—including learning (Baddeley & Hitch, 1974).

Visual memory provides a fruitful domain in which to investigate the impact of internal attention on long-term learning. Specifically, what constitutes a unit of visual memory remains contentious (Fougnie et al., 2010), including theories based on: objects (Luck & Vogel, 1997), feature dimensions (Olson & Jiang, 2002, Wheeler & Treisman, 2002), and in-

formation load (Alvarez & Cavanagh, 2004, Bays & Husain, 2008, Brady et al., 2009). All of these theories share an assumption that the contents of memory are solely determined by properties of stimuli in the current display. However, we propose that the lack of consensus partly results from neglecting the role of prior experience in shaping how a given display or object is construed. In particular, beyond examining how prior perceptual experience affects visual memory (Brady et al., 2009, Curby et al., 2009), we consider how prior retrieval experience tunes object representations. Initial insights can be gleaned from a version of the standard retrieval-induced forgetting task that used visual stimuli (Ciranni & Shimamura, 1999). Instead of relying on semantic associations between items, this study manipulated the perceptual similarity of items. Participants learned the location of twelve objects that were grouped by shape (e.g., circles, triangles, crosses) or color (e.g., green, orange, purple), practiced retrieving an associated feature of a subset of items, and were then worse at recalling features of unpracticed items that were perceptually grouped with practiced items. These findings suggest the presence of competition between objects in visual memory.

1.0.4 Current Study

Here we investigate the impact of internal attention on object learning by examining the role of selective retrieval of features from visual short-term memory (VSTM). After viewing a complex object defined along multiple feature dimensions (e.g., color, orientation, location), we asked: How does selection of one feature from its representation in short-term memory affect how this object is represented in long-term memory? We aim to address two key aspects of visual object learning in the current study: First, motivated by the modal model of memory, we test how accessing an object in a short-term store influences how that

object comes to be represented in a long-term store. Second, building on extant theories of visual memory, we explore how retrieval history influences the weighting of features within individual objects. These goals are distinct from prior work on retrieval-induced forgetting, for example, which examined the mnemonic consequences of retrieving consolidated representations in long-term memory rather than nascent representations in short-term memory, and which examined competition between items that are semantically or visually related rather between features within items.

To accomplish these goals, we designed a visual memory task that manipulates feature retrieval history on an item-level basis (Fig. 1.1). Each novel object was characterized by a unique: angular location relative to central fixation, orientation defined as degree of rotation about center of the shape, and color chosen from a perceptually uniform space. On each trial of a practice phase, observers viewed an object in the periphery of the display. After offset, observers were prompted by a post-cue to report one of its features from VSTM by adjusting a memory probe; a post-cue was used to manipulate internal, rather than external, attention. Observers adjusted the probe by moving the mouse cursor along the wheel until it matched the original object on the cued dimension. All objects were viewed an equal number of times during the practice phase.

To measure the strength of visual long-term memory (VLTM) for the features of each object, observers completed a final test phase with only a feature cue and memory probe. An advantage of using location, orientation, and color as features is that these dimensions are continuous and can be mapped onto a circular space. This approach affords precise measurement of the fidelity of a feature representation, by computing the angular deviation of responses from the true value on the circle (Zhang and Luck, 2008).

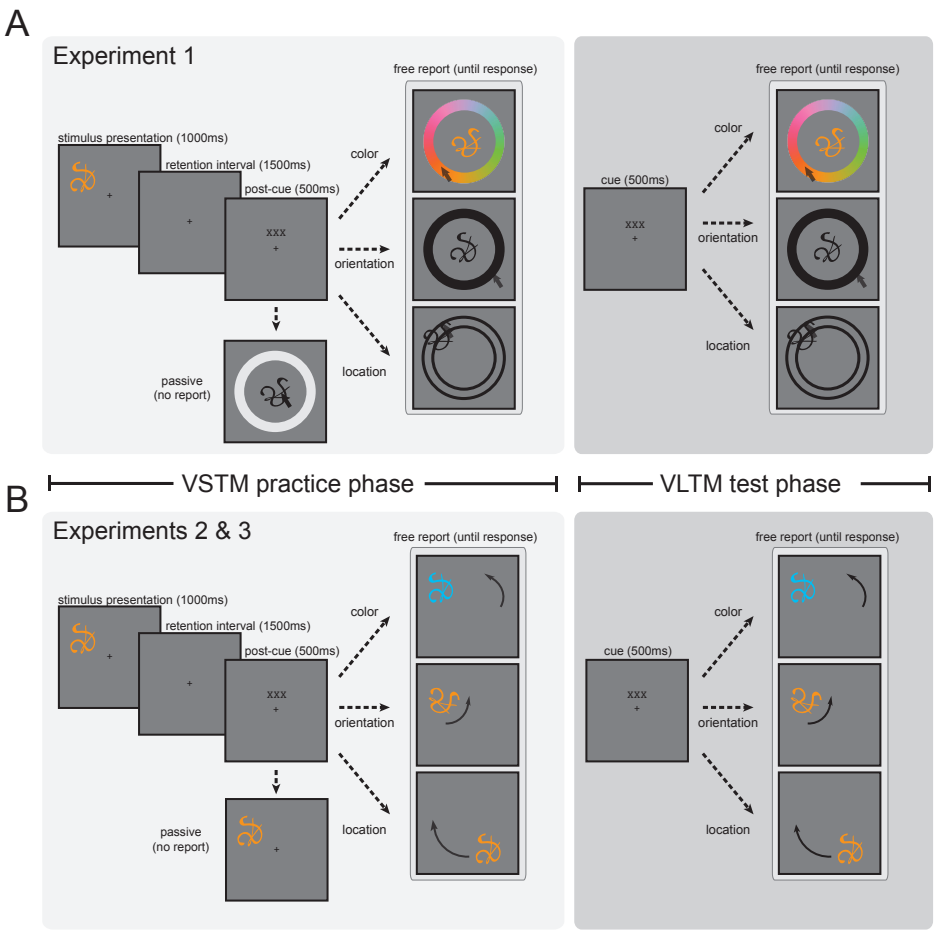


Figure 1.1: Task display. On each trial of the practice phase, observers saw an object defined by a unique color, orientation, and location. After offset, observers were prompted by a post-cue ("xxx" replaced by "color", "orientation", "location", or "click") to report one of its features from VSTM. Observers reported this feature by continuously adjusting a memory probe until it matched the original stimulus. All objects were viewed multiple times during the practice phase. This was followed by a test phase in which observers could rely only on VLTM to report the features of a probe. The nature of the probe differed across experiments. (A) In Experiment 1, the memory probe was a black, canonically oriented version of an object presented in the center of the display. Observers used a visible response wheel to manipulate the memory probe along the cued dimension. (B) In Experiments 2 and 3, the memory probe was identical to the original object in appearance, except for the cued dimension. Observers adjusted the probe by moving the mouse until a good match was found. There was no response wheel, and the mapping between mouse position and feature space was perturbed from trial-to-trial.

We conducted three experiments with different groups of observers. In Experiment 1, we explored the relationship between repeated retrieval of a novel object's features from VSTM and subsequent VLTM for this object. Experiment 2 eliminated alternative explanations of our results based on the amount of sensory experience and the possibility for motor learning. Experiment 3 explored the impact of retrieving multiple features from VSTM on VLTM. Together, these studies reveal important consequences of selective internal attention to features on the organization of memory for objects.

1.1 Experiment 1: Retrieval History and the Tuning of Object Representations

This experiment serves as an initial exploration of the benefits and costs of retrieval from VSTM on VLTM. Objects were assigned to one of three conditions (Figure 1.2A): In the Constant and Switch conditions, the same feature was probed throughout the practice phase; in the test phase, the same feature or an unpracticed feature was probed, respectively. In the Passive condition, objects were studied the same number of times in the practice phase, but no features were reported; in the test phase, an unpracticed features was probed, providing a baseline measure of feature memory in the absence of retrieval history.

If selection of one feature of an object from VSTM just enhances the representation of that feature in VLTM (practice-benefit hypothesis), then test accuracy should be higher for Constant than Passive, and Switch and Passive should not differ. Alternatively, if selecting a feature enhances the representation of all features (object-based hypothesis), then both Constant and Switch should have higher accuracy than Passive. Finally, if this selective operation induces competition among features vying for retrieval (feature-competition hypothesis), then test accuracy should be higher for Constant than Passive and lower for Switch than

Passive. If VLTM is unaffected by accessing VSTM, all conditions should produce quantitatively similar performance; note that this null hypothesis does not only rely on null effects, as test performance in all conditions should be better than chance.

1.1.1 Methods

Participants

Thirty-six naïve observers (23 women, mean age 19.8y) participated in this experiment. In all experiments, observers received course credit or $12 for participating, provided informed consent, and reported normal or corrected-to-normal visual acuity and color vision. The study protocol was approved by the Princeton IRB.

Stimuli

Eight 'alphabets' containing eighteen shapes each were used. Shapes were selected from an online repository of non-Roman ideograms (http://symbols.com, HME Publishing) and reproduced freehand in a vector graphics drawing program. Shapes in an alphabet tended to share common contours, but were otherwise perceptually distinct, as well as rotationally and reflectively asymmetrical. We chose these shapes as stimuli because they were highly novel, preventing prior experience from contaminating our investigations of object learning. Each shape served as the base for one object, with which values from three clearly defined feature dimensions (location, orientation, and color) were associated to form more complex objects. This general approach has been used previously in the study of how objects are represented in VSTM (e.g., Luck & Vogel, 1997). Each shape was displayed at a fixed eccentricity (8°), and was assigned a single angular location (relative to central fixation), orientation (degree of

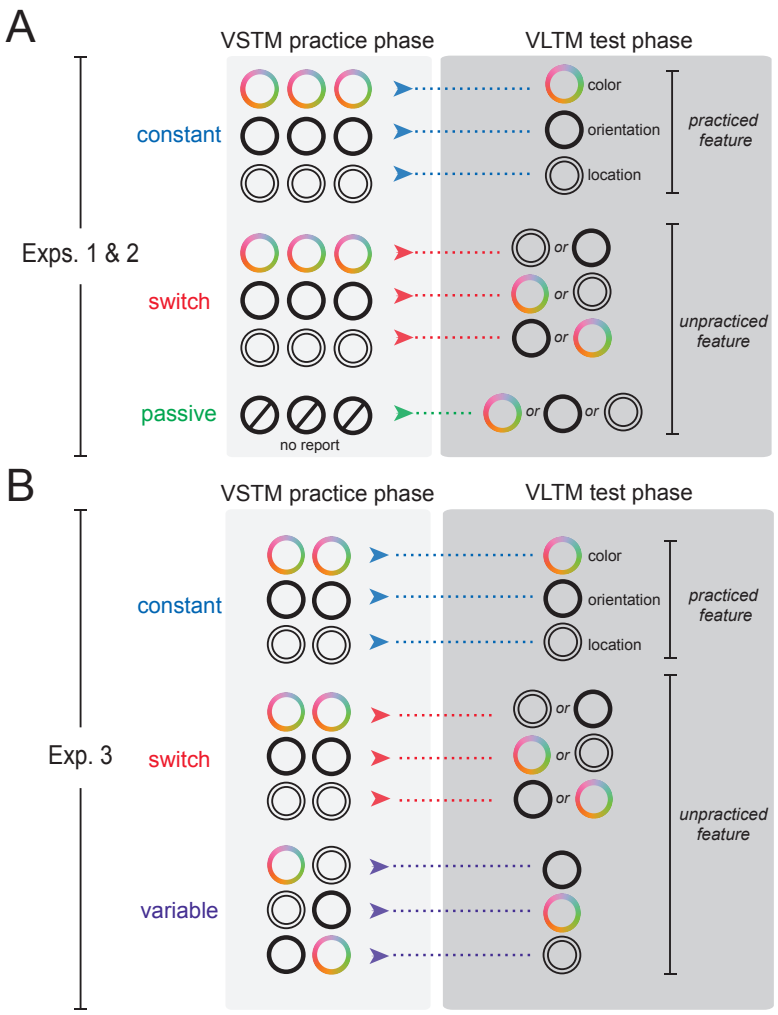


Figure 1.2: (A) Design. In Experiments 1 and 2, all objects were viewed three times during the practice phase. Each object was assigned to the Constant, Switch, or Passive condition. For Constant objects, the same feature was reported throughout both the practice and test phases. Switch objects were probed on the same feature in the practice phase but on an unpracticed feature in the test phase. Passive objects were viewed the same number of times during the practice phase, but no features were reported; one of the unpracticed features was subsequently tested. (B) In Experiment 3, all objects were viewed twice during the practice phase. Each object was assigned to the Constant, Switch, or Variable condition. Constant and Switch conditions were identical to before (but with two repetitions). For Variable objects, two different features were retrieved in the practice phase, and like the Switch condition, an unpracticed feature was subsequently tested.

rotation about center of shape relative to arbitrary canonical orientation), and color (from a perceptually uniform space: equiluminant CIE L*a*b* color space centered at L = 70, a = 20, b = 38; radius 60). Feature values in each dimension were independently and randomly sampled from 180 circular coordinates. Within each alphabet, no two objects could exhibit the same combination of feature values. Stimuli were presented using MATLAB and Psych-Toolbox (Brainard, 1997, Pelli, 1997).

Procedure

The experiment consisted of eight blocks, and each block was randomly assigned a unique stimulus alphabet. Within each block, observers completed two phases: a VSTM practice phase followed by a VLTM test phase. Before starting the experiment, observers were briefed on the two-phase structure of each block and instructed to aim for high accuracy in both practice and test phases of all blocks. A repeated practice-test block design was used instead of a single practice-test session to minimize fatigue and proactive interference during test. That is, while a longer interval between study and test should not have affected VLTM much, we were concerned that these other sources of noise might contaminate our primary dependent measure. Within a block, each of the eighteen objects in the alphabet was presented three times during the practice phase and once during the test phase. The order in which objects appeared was randomized, with the constraint that in the practice phase, every object had to appear at least once before any other object could repeat. Six objects were assigned to each of the Constant, Switch, and Passive conditions within-subject, and these assignments were counterbalanced across subjects. In the Constant and Switch conditions, the same feature was probed three times in the practice phase, and either this

practiced feature (Constant) or an unpracticed feature (Switch) was probed in the test phase. In the Passive condition, objects were presented three times but no features were reported in the practice phase, and an unpracticed feature was probed in the test phase. Across the objects in every condition, the assignment of dimensions to be probed in the practice and test phases was carefully counterbalanced. In each block there were 54 trials in the practice phase (6 objects x 3 conditions x 3 presentations) and 18 trials in the test phase (6 objects x 3 conditions x 1 presentation). Across blocks, there were a total of 432 VSTM practice trials and 144 VLTM test trials. The full procedure lasted less than an hour with instructions and debriefing. On each trial of the practice phase, observers were presented with a single object for 1000 ms under free-viewing conditions. Following its offset and a retention interval lasting 1500 ms, observers were prompted by a text post-cue to report one of the object's features from VSTM. Observers freely adjusted this feature of the memory probe—an initially black, canonically oriented, and centered version of the shape—until it matched the original object on the cued dimension. A continuous response wheel surrounded the memory probe, which observers clicked on to select a value: for the color cue, this entailed choosing from a color wheel; for the orientation cue, this entailed clicking the response wheel at the angle that reflected the desired rotation; for the location cue, this entailed clicking the spatial position of the original stimulus. To make Passive trials as similar as possible to the Constant and Switch conditions (which required responses), observers were presented with the memory probe and a blank response wheel, and were instructed to click on the probe itself to proceed. On trials requiring reporting of a feature (Constant/Switch), the mouse cursor was initially positioned at the center of the display, and the observers moved it to the wheel and clicked. On trials not requiring a report (Passive), the mouse cursor was initially

positioned at a random angle on the response wheel, and observers moved it to the center and clicked. This central movement in the Passive condition was chosen to ensure that observers attended to the memory probe, which they could have otherwise ignored on these trials. Nevertheless, the magnitude of the movement was equated across conditions. On test phase trials, only the text cue and memory probe were presented, requiring observers to report cued features purely from VLTM. Presenting one object per trial helped limit possible feature interactions to those between features within an object. A downside of this approach is that observers might have had enough spare resources to represent each feature separately rather than as part of a bound object. However, we felt that the alternative—presenting multiple objects per trial—would have been less ideal for three reasons: (1) simultaneous presentation of multiple objects would increase the likelihood of interactions between features from different objects, obscuring interactions between features from the same object; (2) the binding of features into a coherent object requires focused attention (Treisman, 2006), and a set size of one ensured that the object would be fully attended; and (3) as set size increases, observers may default to summary statistics over feature dimensions, which can bias individual object representations (Brady & Alvarez, 2011). Nevertheless, future research should examine the consequences of selective feature retrieval from VSTM for larger set sizes.

Measurement of memory performance

Response error was calculated as the angular deviation from the position on the wheel corresponding to the correct value on the cued dimension. Because responses were continuous along each feature dimension, this measure permitted quantitative estimates of the error distributions for each condition. Narrower distributions reflected more accurate performance,

with more responses clustered about the true value. To quantify accuracy in each condition, the root average squared deviation from the true value in degrees was analyzed (root mean squared error, RMSE). RMSE is a standard estimation approach that aggregates errors across individual observations into a single measure of performance and is intuitively expressed in the same units being estimated (in this case, degrees). When observers' responses are unbiased, RMSE is formally equivalent to the standard deviation of errors. There are other prominent approaches for characterizing performance in continuous report tasks, including fitting a mixture of theoretical distributions using maximum likelihood estimation (Zhang & Luck, 2008, Wilken & Ma, 2004). Although we employ this mixture modeling technique later to explore different components of performance related to "accessibility" and "precision," we focus primarily on RMSE because it provides a more direct measure of performance, requiring fewer assumptions about the underlying response distributions and coping better with the relatively small number of trials we could obtain in the VLTM test. We modeled chance performance as a uniform distribution over the interval [-180°, 180°) corresponding to the range of possible errors on the report task, where 0° represents the correct value. The expected magnitude of errors made by an observer guessing randomly and with no information about the correct feature value would be 90°. However, the RMSE under pure guessing conditions is given by the expression (360°)/□12□ 103.9° (Rinaman et al., 2012), derived from the definition of standard deviation of a uniform distribution (see dotted line in Figure 1.3B). The height of its probability density function is 1/(360°) = 0.0028 (see dotted line in Figure 1.3A). In analyzing RMSE, we collapsed over the feature dimension being probed within each condition. Because the design was carefully counterbalanced such that each feature was equally likely to be practiced and tested, there were many feature permuta-

tions and thus we had insufficient statistical power for examining features separately (some permutations occurred as rarely as once per block). Whether there are feature-specific differences in the consequences of retrieval practice remains an important question for future research. This could be examined by reducing the number of dimensions or probing a single dimension for all objects. Because we were interested in measuring the consequences of retrieval from VSTM on VLTM, only objects whose features were successfully reported during the practice phase were retained for analysis. Objects were excluded from analysis if, on any of the practice phase trials, the RMSE fell 1.5 times the interquartile range (IQR) outside the 1st and 3rd quartiles in each condition (Frigge et al., 1989, Tukey, 1977). This outlier exclusion procedure was independent of VLTM test performance, our primary dependent measure. On average, 1.02 objects per condition and block were excluded from further analyses as outliers; this rate did not differ between the Constant and Switch conditions ($t(35) = 1.25$, $p = 0.\ 219$, $d = 0.209$).

1.1.2 Results and Discussion

In the practice phase, performance was better than chance at all three repetitions in both the Constant and Switch conditions (all $t(35)$s > 225.8, ps $\ll 0.001$, ds > 37.6). We analyzed the practice phase data using a 2 (condition) × 3 (repetition) repeated-measures ANOVA. Differences in accuracy between Constant and Switch were not expected because these conditions were identical until the test phase (Constant RMSE [first, second, third] = 11.35°, 11.05°, 10.63°; Switch RMSE [first, second, third] = 11.11°, 10.83°, 10.72°). Indeed, there was no main effect of condition ($F(1,35) = 0.152$, $p = 0.699$, $\eta p2 = 0.004$). Despite high overall accuracy, there was a main effect of repetition with performance improving over the three repetitions

($F(2,70) = 5.51$, $p = 0.006$, $\eta p2 = 0.136$). There was no interaction between condition and repetition ($F(2,70) = 0.382$, $p = 0.684$, $\eta p2 = 0.011$). We hypothesized that having reported one feature of an object during the practice phase would, in the test phase, result in better VLTM for that feature (Constant) and worse VLTM for other features of the same object (Switch), relative to baseline (Passive). As can be seen in Figure 1.3A, aggregate test phase errors were consistent with this prediction (Constant RMSE = 55.60°; Switch RMSE = 75.22°; Passive RMSE = 70.37°). To assess the reliability of these differences between conditions, we compared RMSE across observers (Figure 1.3B). Error was reliably lower for Constant than Passive ($t(35) = 6.31$, $p < 0.001$, $d = 1.05$) and higher for Switch than Passive ($t(35) = 2.24$, $p = 0.031$, $d = 0.374$). This latter difference is striking because both Switch and Passive test trials probed unpracticed features, differing only with respect to retrieval history of other features. Performance was better than chance in all three conditions (all $t(35)$s > 10.4, $p \ll 0.001$, ds > 1.73). These results suggest that selecting one feature of a novel object from VSTM biases the resultant representation towards the attended feature and away from unattended features.

A potential explanation for why VLTM performance was worse in the Switch condition is that there was response-level interference between the practiced feature and unpracticed features, especially for features with a spatial component (e.g., location and orientation). For instance, making a motor movement to report the color of an object might interfere more with memory for the object's location than reporting the location of an object would interfere with memory for the object's color. A key prediction of this account is that the degree of VLTM impairment for an unpracticed feature (i.e., the cost for Switch relative to Passive) will differ between objects where a non-spatial feature (i.e., color) was practiced but a spatial feature (i.e., orientation or location) was tested vs. where a spatial feature was practiced

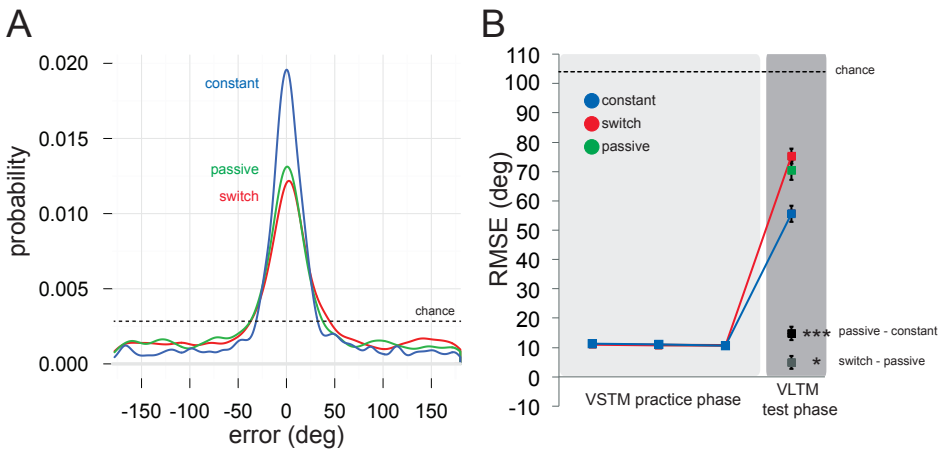


Figure 1.3: Results of Experiment 1. (A) Empirical distributions depicting the probability of observing an error of a given magnitude in the VLTM test for each condition. These distributions were constructed for visualization purposes only, by pooling data across all observers. (B) Task performance across VSTM practice and VLTM test phases as measured by RMSE for objects successfully reported during practice phase. In the test phase, practiced features (Constant) were reported most accurately, while unpracticed features of practiced objects (Switch) were impaired relative to unpracticed features of unpracticed objects (Passive). The height of the dotted line represents expected performance under random guessing. Error bars reflect ±1 SEM. *p<0.05, ***p<0.001.

but a non-spatial feature was tested. To examine this possibility, we performed a 2 (feature type: spatial vs. non-spatial) x 2 (condition: Switch vs. Passive) repeated-measures ANOVA on VLTM performance. The main effect of condition remained, with Switch impaired relative to Passive ($F(1,35) = 4.22$, $p = 0.0475$, $\eta p2 = 0.108$; p-value differs from above because only a subset of the feature transitions were included). However, inconsistent with the possibility that unpracticed spatial features suffered more interference, there was no interaction between feature type and condition ($F(1,35) = 0.0173$, $p = 0.896$, $\eta p2 < 0.001$). Directing internal attention to information in short-term memory has been shown to enhance subsequent memory (Johnson et al., 2002). Our results build on these findings by showing that diverting internal attention away from information can also impair its representation in long-term memory. That is, we interpret our results as evidence that deploying internal attention concurrently enhances selected features while suppressing other unselected features. This tuning may result from an inhibitory mechanism needed to resolve competition between representations of the cued and uncued features of the probe that are simultaneously active in VSTM. This kind of explanation has been applied to memory for arbitrary lexical and semantic associations (Carpenter et al., 2006, Carrier & Pashler, 2002, Levy & Anderson, 2002), but here we demonstrate a novel role for feature-level inhibition during object learning.

1.2 Experiment 2: Ruling Out Alternative Explanations

This experiment seeks to rule out two alternative explanations for the difference between Switch and Passive in Experiment 1. First, although Switch and Passive conditions were equated in testing an unpracticed feature, they differed in that observers received extra

sensory exposure to the cued feature of Switch objects while it was reported in the practice phase. That is, additional encoding (rather than retrieval) of the practiced feature may have strengthened the weight of that feature relative to the unpracticed test feature. Second, observers may have associated Switch objects with the location on the response wheel that was needed to report the cued feature in the practice phase. Because the same wheel was used for all feature dimensions, such response learning could have caused interference when the unpracticed test feature required a different response. A related issue is that the presence of these spatial landmarks and repeated motor trajectories may have inherently interfered with the VSTM representation of unpracticed features with a spatial component, such as location and orientation. To remove these potential confounds, we modified the task from Experiment 1 by changing how the memory probe was presented and how observers made responses. To equate sensory exposure across dimensions, the memory probe—rather than being black, canonically oriented, and centered—was identical to the original object except for the cued dimension, which was set to a random starting value. For example, if post-cued to report color, the memory probe appeared in the same location and orientation as the object, but in a random color. To eliminate the possibility of response learning, the mouse cursor and wheel were removed from the display. (Responses were tracked on a hidden wheel, but observers could only base their response on changes in the memory probe.) Moreover, the mapping between cursor position and feature space was randomly rotated on every trial to ensure that novel mouse trajectories were always required to arrive at any given feature value. If the results of Experiment 1 reflect response learning, then differences between conditions should be eliminated by the lack of a consistent response in the practice phase. If the results reflect biased sensory exposure, then accuracy should be markedly different than Ex-

periment 1, with Switch similar to Passive, and both better than Constant. Consider the example object above: While reporting color in the practice phase, there was greater exposure to the unpracticed location and orientation features because these values were always veridical in the probe; in contrast, the practiced color feature was initially random and only close to the "true" value near the time of response. Thus, if this object was in Constant condition, the amount of exposure to the test feature (color) would be less than if it was in the Switch (veridical location or orientation) or Passive (all features veridical). Alternatively, according to our hypothesis that selecting a feature in VSTM suppresses other unpracticed features in VLTM, we should replicate the lower accuracy for Switch relative to Passive despite equivalent sensory exposure to unpracticed features in both conditions. We did not have strong predictions for Passive vs. Constant, since a benefit of repeated testing for Constant could be diluted by a benefit of greater sensory exposure for Passive.

1.2.1 Methods

Participants

Thirty naïve observers (17 women, mean age 20.3y) participated in this experiment.

Procedure

The stimuli and procedure were identical to Experiment 1, with four changes (Fig.1.1b): (1) observers adjusted the mouse position to report the cued feature of the memory probe, but other visual feedback was eliminated (i.e., no response wheel or mouse cursor); (2) the mapping between the (hidden) cursor position and feature space was randomly rotated from trial-to-trial to ensure a unique mouse trajectory each time an object was probed; (3) the

memory probe was identical to the original object in appearance, except for the cued dimension, which was set to a random starting value on each trial; (4) in the Passive condition, where no dimension was cued in the practice phase, the original object was presented intact for an interval that matched the running mean response time of all Constant and Switch trials in that block—this delay was used to ensure equal average sensory exposure to all objects. Because this experiment lasted longer, observers completed five practice-test blocks, with alphabets sampled randomly from the original corpus of eight alphabets.

1.2.2 Results and Discussion

In the practice phase, performance was better than chance at all three repetitions in both the Constant and Switch conditions (all $t(29)$s > 151.0, ps ≪ 0.001, ds > 27.6). On average, 1.19 objects per condition and block were excluded from further analyses as outliers; this rate did not differ between Constant and Switch ($t(29) = 0.359$, $p = 0.722$, $d = 0.0656$). We analyzed the practice phase data using a 2 (condition) × 3 (repetition) repeated-measures ANOVA. Accuracy did not differ between Constant and Switch ($F(1,29) = 0.066$, $p = 0.799$, $\eta p2 = 0.002$). Unlike Experiment 1, the improvement across repetitions did not reach significance ($F(2,58) = 2.10$, $p = 0.132$, $\eta p2 = 0.067$), although the range of performance was comparable (Constant RMSE [first, second, third] = 10.68°, 10.04°, 10.08°; Switch RMSE [first, second, third] = 10.71°, 10.28°, 10.19°). This suggests that observers may have been performing near ceiling in the VSTM task in this experiment. Regardless, retrieval practice can affect later test memory even when performance is at ceiling (Roediger & Karpicke, 2006). There was no interaction between condition and repetition ($F(2,58) = 0.074$, $p = 0.929$, $\eta p2 = 0.003$).

Consistent with our hypothesis, the distribution of errors pooled across observers mir-

rored that of Experiment 1 (see Figure 1.4A; Constant RMSE = 56.77°; Passive RMSE = 59.69°; Switch RMSE = 68.86°). Assessing the reliability of these differences across observers (Figure 4B): the decrement for Switch vs. Passive again reached significance ($t(29) = 2.77$, $p = 0.0096$, $d = 0.506$), but the advantage for Constant vs. Passive did not ($t(29) = 1.00$, $p = 0.326$, $d = 0.182$); Constant was nevertheless more accurate than Switch ($t(29) = 5.19$, $p < 0.001$, $d = 0.948$). Finally, performance was better than chance in all conditions (all $t(29)$s > 10.4, ps $\ll 0.001$, ds > 1.91). We placed stress on the hypothesized role of selective feature retrieval in tuning memory for objects by controlling the amount of sensory exposure and eliminating the possibility of response learning. The results suggest that suppression of VLTM for unpracticed features is a robust consequence of selection within VSTM, even when those features are externally available during this internal selection.

Although the overall pattern of results matched Experiment 1, the benefit of retrieval practice for Constant vs. Passive was mitigated. We interpret this smaller effect as the canceling out of a practice benefit for Constant (that we observed in Experiment 1) by additional encoding for Passive (which was not part of Experiment 1). That is, in the Passive condition, all three features were fully present on the memory probe, whereas in the Constant condition, observers needed to regenerate the practiced and later tested feature, moving through intermediate values that may have diluted the representation. Consistent with this interpretation, accuracy in the Constant condition of both experiments was nearly identical (55.60° in Experiment 1 vs. 56.77° in Experiment 2; $t(64) = 0.30$, $p = 0.767$, $d = 0.0737$). What differed between experiments was the Passive condition (70.37° in Experiment 1 vs. 59.69° in Experiment 2; $t(64) = 2.18$, $p = 0.033$, $d = 0.538$), as would be expected if the smaller effect was caused by greater sensory exposure (which only affected unpracticed features). Never-

theless, if the enhancement effect were stronger, it would have overcome this effect of sensory exposure and revealed itself again. For now, the enhancement effect in Experiment 1, while robust, awaits further study and validation (see also Section 5 for additional tentative evidence from estimation of guessing rates).

Another interesting trend emerged when comparing Experiments 1 and 2. Namely, in addition to the reliable difference between Switch and Passive, Switch did not benefit as much from greater sensory exposure as Passive (75.22° in Experiment 1 vs. 68.86° in Experiment 2; $t(64) = 1.51$, $p = 0.136$, $d = 0.373$). In fact, the size of the Switch vs. Passive difference roughly doubled in magnitude. Although this is a post-hoc finding that requires further investigation, one intriguing possibility is that the correspondence between the externally available features of the probe and the representation of the object in VSTM may have externalized and strengthened the competition—otherwise internally mediated—between active feature representations, leading to greater weakening. Indeed, stimuli that match the contents of working memory automatically capture attention (Soto et al., 2008a). Relatedly, the external availability of unpracticed features may have allowed them to seep into VSTM, resulting in even greater inhibition when the cued feature was attended and reported.

1.3 Experiment 3: Variable vs. Focused Retrieval History

In Experiments 1 and 2, we examined the impact of selective retrieval of one feature from VSTM on object learning. However, it is not uncommon for more than one feature of an object to be relevant over repeated encounters. This experiment tests how selection of multiple features from VSTM affects VLTM for novel objects. Objects were assigned to one of three conditions (Figure 1.2B): In the Constant and Switch conditions, the same feature was

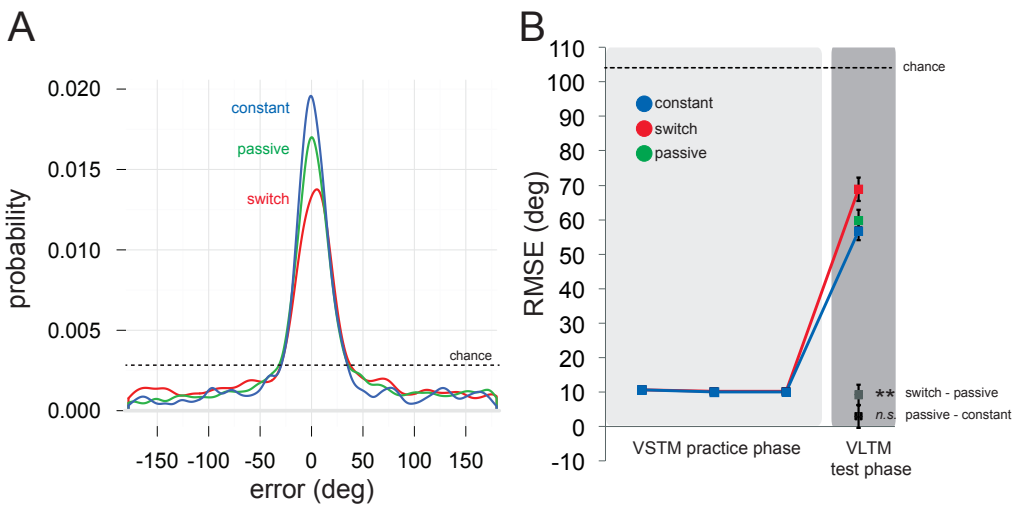


Figure 1.4: Results of Experiment 2. (A) Empirical distributions for each condition in the VLTM test phase, pooling data across observers for visualization purposes. (B) Task performance across VSTM practice and VLTM test phases as measured by RMSE. In the test phase, unpracticed features of practiced objects were impaired (Switch), even when observers received equal (vs. Passive) or greater (vs. Constant) veridical sensory exposure to these features during the practice phase. Error bars reflect ±1 SEM. ⍰⍰$p < 0.01$.

probed twice in the practice phase; in the test phase, the same feature or an unpracticed feature was probed, respectively. In the Variable condition, two different features were probed in the practice phase; in the test phase, the third (unpracticed) feature was probed. This design affords a direct comparison between variable practice (Variable) and focused practice (Switch) with respect to long-term memory for unpracticed features. There were three possible outcomes: First, variable practice might moderate the suppression of unpracticed features by releasing inhibition when previously unpracticed features must be practiced, resulting in higher accuracy in Variable than Switch. Second, the distinction between variable and focused practice may be inconsequential insofar as all that matters for unpracticed features is that they are not selected, in which case Variable and Switch accuracy would not differ (but should be better than chance). Third, variable practice might induce greater suppression of always unpracticed features due to the weak status of previously unpracticed features that must be practiced, which incites strong competition and results in higher accuracy for Switch than Variable.

1.3.1 Methods

Participants

Thirty naïve observers (18 women, mean age 20.0y) participated in this experiment.

Procedure

All procedures and stimuli were identical to Experiment 2, except that each object was presented twice during the practice phase and the Passive condition was replaced by a Variable condition in which two different features were practiced and the third (unpracticed) feature

was tested.

1.3.2 Results and Discussion

In the practice phase, performance was better than chance for both repetitions in the Constant, Switch, and Variable conditions (all $t(29)$s > 138.2, ps « 0.001, ds > 25.2). On average, 0.90 objects per condition and block were excluded from further analyses as outliers; this rate did not differ between Constant and Switch ($t(29) = 1.66$, $p = 0.107$, $d = 0.304$) or between Switch and Variable ($t(29) = 1.30$, $p = 0.204$, $d = 0.237$). However, 0.23 more objects were excluded per block from Variable than Constant ($t(29) = 3.12$, $p = 0.0041$, $d = 0.569$). This difference may be attributable to the fact that the Variable condition involved practicing two features. However, since Constant and Switch conditions were identically structured during this phase, and the Switch exclusion rate did not differ from that of Variable, the Constant/Variable difference may not be meaningful. Regardless, additional analyses verified that the exclusion procedure did not affect the pattern of VLTM results. We analyzed the practice phase data using a 3 (condition) × 2 (repetition) repeated-measures ANOVA. Performance did not differ across conditions ($F(2,58) = 1.01$, $p = 0.372$, , $\eta p2 = 0.034$), despite the fact that in the second presentation the same feature was practiced for Constant and Switch and a new feature was practiced for Variable. Accuracy did not differ across the two repetitions ($F(1,29) = 1.08$, $p = 0.308$, $\eta p2 = 0.036$; Constant RMSE [first, second] = 11.13°, 10.54°; Switch RMSE [first, second] = 10.37°, 10.37°; Variable RMSE [first, second] = 10.91°, 10.96°). There was no interaction between condition and repetition ($F(2,58) = 0.788$, $p = 0.460$, $\eta p2 = 0.026$). Consistent with the other experiments (A), VLTM was better for practiced features (Constant RMSE = 60.10°) than unpracticed features in cases

where one other feature had been practiced (Switch RMSE = 68.35°). Practicing two features led to even worse VLTM for unpracticed features (Variable RMSE = 75.07°). Assessing the reliability of these differences across observers (Figure 1.5B): the decrement for Switch vs. Constant reached significance ($t(29) = 2.77$, $p = 0.0096$, $d = 0.506$), as did the decrement for Variable vs. Switch ($t(29) = 3.08$, $p = 0.0045$, $d = 0.563$). Finally, performance was better than chance in all conditions (all $t(29)$s > 9.66, ps $\ll 0.001$, ds > 1.76).

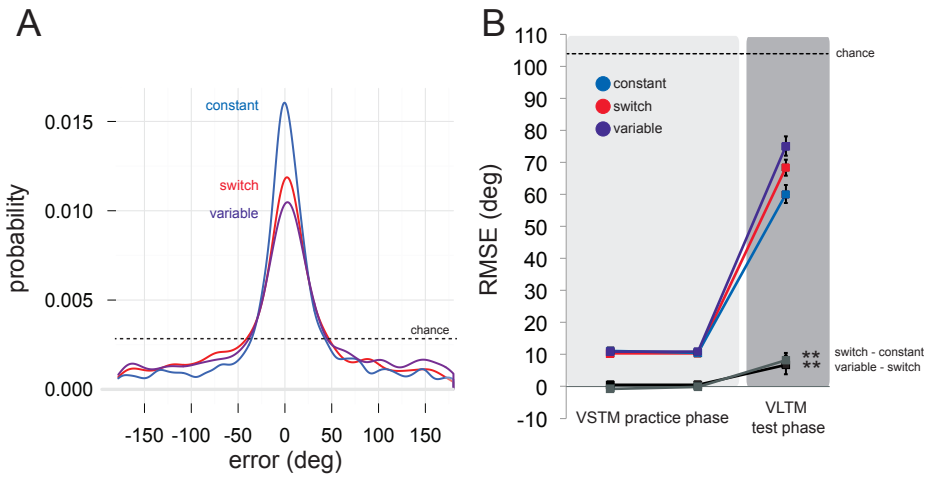


Figure 1.5: Results of Experiment 3. (A) Empirical distributions for each condition in the VLTM test phase, pooling data across observers for visualization purposes. (B) Task performance across VSTM practice and VLTM test phases as measured by RMSE. In the test phase, unpracticed features were impaired when belonging to an object for which two (Variable) vs. one (Switch) other features had been practiced. Error bars reflect ±1 SEM. **p<0.01.

Despite equating study and retrieval opportunities, selecting multiple features of an object from VSTM suppressed VLTM for other features more than selecting a single feature repeatedly. Indeed, under both variable and focused practice, an unpracticed feature was finally tested—the objects differed only with respect to the nature of prior retrieval practice. These findings contradict the simplest conception of retrieval tuning, whereby only the

number of retrieval attempts predicts the strength of subsequent memory for practiced and unpracticed features. Moreover, lower accuracy under variable practice stands in contrast to other examples of beneficial variability, such as 'double-training' in perceptual learning (Xiao et al., 2008) and the 'spacing effect' in episodic memory (Melton, 1970). Instead, these results are reminiscent of the finding that 'extinction' is stronger under variability: When a conditioned stimulus is not reinforced in multiple contexts vs. a single context, then subsequent renewal of its association with an unconditioned stimulus is attenuated (Chelonis, Calton, Hart, & Schachtman, 1999). Although a speculative connection at this stage, to the extent that the color, orientation, and location reports served as distinct task contexts, an unreported feature passed over once in two such contexts may be more strongly inhibited than when neglected twice in the same context.

1.4 Modeling Accessibility and Precision

What is the nature of the observed changes in VLTM in Experiments 1-3? One possibility is that retrieval practice of one feature dimension enhances the accessibility of representations in that dimension relative to other unpracticed dimensions. Another possibility is that the observed memory effects can be explained by changes in the precision of representations in the practiced vs. unpracticed dimensions. To address this issue quantitatively, we estimated the relative contributions of decrements in accessibility and precision using a model-fitting method that partitioned errors in the final test phase into two component distributions reflecting: the probability of guessing (i.e., inversely related to accessibility) and the precision of retrieved feature representations.

1.4.1 Methods

On some proportion of trials during the test phase, response errors were very large. Such errors might reflect either a random guess or a very imprecise memory representation. The data can thus be described as a mixture of two distributions: a uniform distribution when the observer is unable to access the feature value and attempts to randomly guess from the full space of feature values, and a von Mises (circular normal) distribution when the observer could access the representation in memory and made a noisy response centered around the true value. Maximum likelihood estimation can be used to recover the likelihood that responses were drawn from each of these distributions by fitting parameters that characterize each distribution (Zhang & Luck, 2008). Two parameters were of primary interest: (1) the height of the uniform component of the mixture, or the guessing rate (g); and (2) the standard deviation of the von Mises component (sd), or the inverse of memory precision. We note one important caveat about our application of this technique: mixture modeling requires assumptions about the underlying distributions that are difficult to validate with a small number of observations. Specifically, we obtained fewer observations per observer and condition (48 in Experiment 1, 30 in Experiments 2 and 3) than previous studies that have used this technique (e.g., 150 in Zhang & Luck, 2008). To help compensate for this, we used a maximum likelihood estimation method in which we repeatedly estimated the best-fitting parameters from random starting values, and then extracted the modal value from this distribution. This iterative procedure was used to ensure that estimates of g and sd were as robust as possible (Figure 1.6A).

1.4.2 Results

Experiment 1

The model fitting procedure suggested that guessing occurred on 30.9% of test trials for Constant (95% CI for g: [25.4%, 36.3%]), 50.0% of trials for Switch ([43.0%, 56.2%]), and 48.2% of trials for Passive ([40.5%, 55.6%]). Guessing was less common for Constant than Switch ($t(35) = 7.04$, $p < 0.001$, $d = 1.17$) and Passive ($t(35) = 6.36$, $p < 0.001$, $d = 1.06$), which did not differ from each other ($t(35) = 0.646$, $p = 0.523$, $d = 0.108$). The precision of non-guess responses was similar across conditions: 15.9° for Constant (95% CI for sd: [14.1°, 17.8°]), 17.9° for Switch ([15.3°, 20.6°]), and 17.7° for Passive ([15.6°, 19.7°]). Indeed, Constant did not differ from Passive ($t(35) = 1.34$, $p = 0.189$, $d = 0.224$) or Switch ($t(35) = 1.24$, $p = 0.223$, $d = 0.207$), and they did not differ from each other ($t(35) = 0.152$, $p = 0.880$, $d = 0.025$). These results suggest that differences in the probability of successful access, rather than the precision of stored representations, may partly underlie the observed changes in VLTM performance due to prior retrieval practice. Individual differences in test phase performance provide further support for this interpretation (Figure 1.6B). RMSE was highly correlated with individual estimates of g (Constant: $r = 0.92$, $p < 0.001$; Switch: $r = 0.90$, $p < 0.001$; Passive: $r = 0.96$, $p < 0.001$), but not sd (Constant: $r = 0.15$, $p = 0.365$; Switch: $r = -0.13$, $p = 0.453$; Passive: $r = 0.006$, $p = 0.972$).

Experiment 2

The modeling results from Experiment 2 followed a similar pattern to Experiment 1. Guessing occurred on 31.3% of test trials for Constant (95% CI for g: [24.8%, 37.8%]), 44.4% of

trials for Switch ([37.7%, 51.5%]), and 37.5% of trials for Passive ([29.6%, 45.4%]). Guessing was less common for Constant than Switch ($t(29) = 4.68$, $p < 0.001$, $d = 0.854$), with a trend for Constant less than Passive ($t(29) = 1.84$, $p = 0.077$, $d = 0.335$). There was also a trend for more guessing on Switch than Passive ($t(29) = 1.86$, $p = 0.073$, $d = 0.339$). The precision of non-guess responses was similar across conditions: 15.3° for Constant (95% CI for sd: [12.7°, 17.9°]), 17.80° for Switch ([13.4°, 22.2°]), and 16.3° for Passive ([14.1°, 18.66°]). No differences between conditions reached significance (ps > 0.335, ds < 0.179). As in Experiment 1, RMSE was highly correlated with individual estimates of g (Constant: $r = 0.85$, $p < 0.001$; Switch: $r = 0.88$, $p < 0.001$; Passive: $r = 0.92$, $p < 0.001$), but generally not with sd (Constant: $r = -0.14$, $p = 0.450$; Passive: $r = 0.16$, $p = 0.400$; Switch: $r = 0.26$, $p = 0.157$).

Experiment 3

Guessing occurred on 34.8% of test trials for Constant (95% CI for g: [27.2%, 42.4%]), 44.9% of trials for Switch ([38.0%, 51.8%]), and 50.6% of trials for Variable ([41.7%, 59.5%]). Guessing was less common for Constant than Switch ($t(29) = 2.66$, $p = 0.01$, $d = 0.487$) and Variable ($t(29) = 3.49$, $p = 0.002$, $d = 0.636$), which did not differ from each other ($t(29) = 1.49$, $p = 0.146$, $d = 0.273$). The precision of non-guess responses was similar across conditions: 22.5° for Constant (95% CI for sd: [16.0°, 29.1°]), 22.4° for Switch ([17.4°, 27.6°]), and 23.1° for Variable ([16.1°, 30.2°]). No differences between conditions reached significance (ps > 0.871, ds < 0.030). RMSE was again highly correlated with individual estimates of g (Constant: $r = 0.88$, $p < 0.001$; Switch: $r = 0.82$, $p < 0.001$; Variable: $r = 0.93$, $p < 0.001$). Conventional analyses suggested a significant, positive relationship between RMSE and sd for Constant ($r = 0.41$, $p = 0.024$), though not Switch ($r = 0.27$, $p = 0.152$) or Variable ($r = -0.08$, $p = 0.654$).

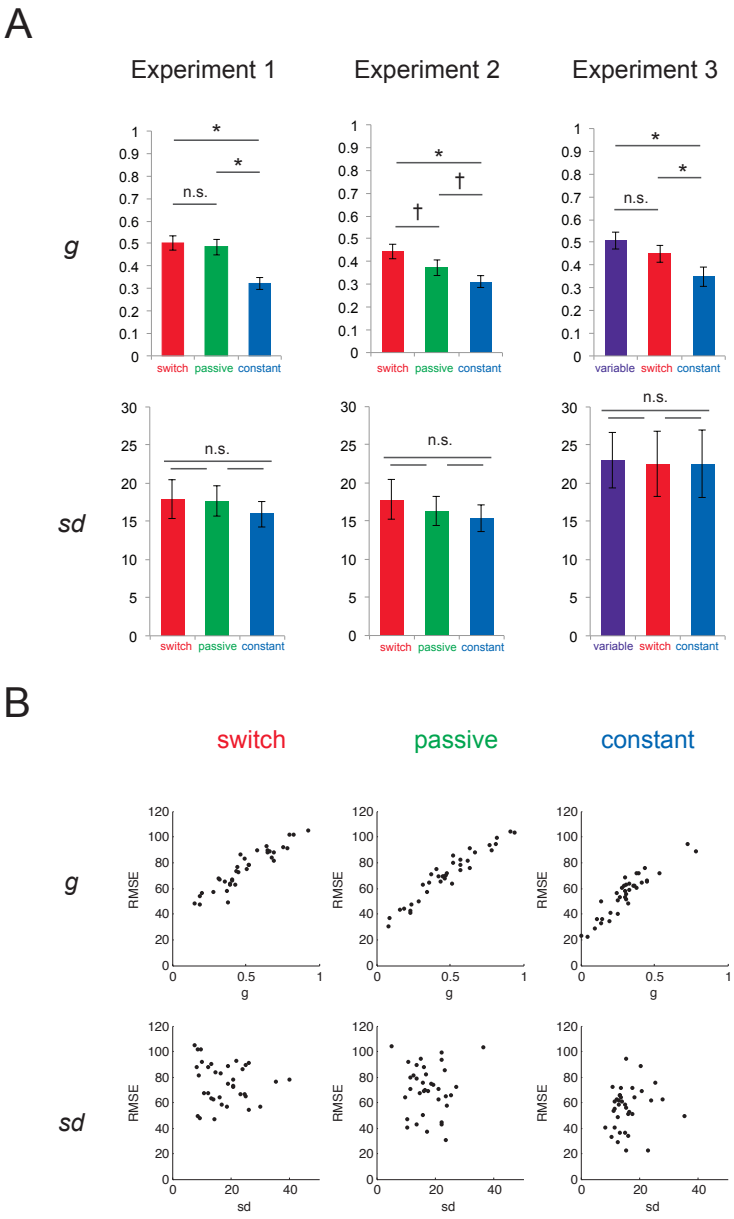


Figure 1.6: Model Fitting Results. (A) Guessing (g) and precision (1/sd) across conditions in the VLTM test phase of each experiment. The probability of guessing was generally reduced for practiced (Constant) vs. unpracticed feature (Switch/Variable/Passive) features, but the precision of non-guesses did not significantly differ between conditions. Error bars reflect ±1 SEM. * $p < 0.05$, † $p < 0.10$. (B) Relationship between parameter estimates and primary RMSE measure in Experiment 1. There were robust positive correlations between RMSE and g, but unreliable correlation between RMSE and sd. This pattern of correlations held across all experiments.

However, inspection of the distribution of sd parameter fits revealed the presence of a single high-leverage observer in the Constant condition (sd = 96.3°, RMSE = 101.8°), whose sd estimate exceeded 3 standard deviations plus the mean sd. Without this observer, the correlation between RMSE and sd for Constant (r = 0.03, p = 0.872) was no longer significant. This result suggests that the apparent relationship between RMSE and sd in for Constant was spurious, likely due to poor parameter estimation in these small samples.

Discussion

The guessing parameter results from all three experiments suggest that retrieving one feature of an object from VSTM affects which features of that object are accessible in VLTM. At the same time, the precision parameter results suggest that VSTM retrieval does not affect the fidelity with which the features of an object are stored in VLTM. Strong correlations between RMSE and guessing but not precision in each condition further suggest that the main RMSE results from all experiments reflect differences in accessibility.

1.5 General Discussion

This study aimed to elucidate how internal attention to features in VSTM influences VLTM for novel objects. Specifically, we explored how retrieval history determines the weighting of features within individual objects over the course of learning. To address this question, we designed a visual memory task that manipulated feature retrieval history on an item-level basis. Observers initially viewed a series of shapes defined by unique values along three feature dimensions (location, orientation, and color), and reported one of these features from VSTM using a continuous report procedure. In a subsequent test phase, observers reported

either the same feature or an unpracticed feature from VLTM. Across three experiments with different groups of observers, we discovered important consequences of internal attention for the organization of object memory. In Experiment 1, we evaluated three hypotheses about how retrieving an object's features from VSTM would affect VLTM for that object: (1) reporting one feature might affect only that feature (practice-benefit hypothesis); (2) internal attention to one feature might benefit all features (object-based hypothesis); and (3) internal attention might induce competitive dynamics among features (feature-competition hypothesis). Consistent with the third hypothesis, for a given object, memory was both enhanced for practiced features and suppressed for concurrently unpracticed features. Retrieval history may thus play an important role in how objects are construed in visual memory, extending views that emphasize only the stimulus properties of the current display (Alvarez & Cavanagh, 2004, Bays & Husain, 2008, Luck & Vogel, 1997, Olson & Jiang, 2002, Wheeler & Treisman, 2002). Moreover, previous work that has manipulated internal attention to items in short-term memory (Lepsien & Nobre, 2006, Nobre et al., 2004b) has focused on immediate consequences. In contrast, we tested the long-term consequences of such orienting for subsequent memory. We found that long-term memory biases for specific items can be acquired rapidly (i.e., after 2-3 exposures) and over a relatively large corpus (i.e., 90-144 novel objects, depending on the experiment). Moreover, although our findings seem to contradict a standard object-based view—that processing one feature of an object benefits processing of its other features—they may be consistent in another sense. Specifically, as expanded below, we interpret our findings as suggesting that all features of an object come to mind during VSTM retrieval in an object-based manner, and that which features are selected vs. inhibited determines how an object is later represented in VLTM. In Experiment

2, we placed greater pressure on the hypothesized role of memory retrieval in shaping the organization of long-term memory for objects by equating the amount of sensory exposure to all feature dimensions, and controlling for motor response learning. While in Experiment 1 the memory probe did not contain any of the three critical features, the memory probe in Experiments 2 and 3 was identical to the original object except for the feature being probed. As a consequence, observers received concurrent veridical access to unpracticed dimensions during the practice phase. Despite greater exposure to unpracticed features, memory for these features remained worse when other features had been practiced (and exposed less). These findings show that the suppression in VLTM is a robust consequence when internal attention neglects features in VSTM, even when these features are externally available. Intriguingly, the size of the Switch vs. Passive difference at test roughly doubled in magnitude in Experiment 2 compared to Experiment 1. One explanation is that the unpracticed features were stronger competitors in Experiment 2 because they were externally available in the probe rather than only internally active in VSTM. This heightened competition may have led to even further weakening of the unpracticed features when the practiced features were successfully reported. Indeed, the unpracticed features of the probe may have automatically captured attention simply because they matched the contents of working memory (Soto et al., 2008a). Just as external attention is recruited to filter out distractors, the perceptual availability of task-irrelevant features may have 'propped up' the unpracticed features in VSTM, requiring stronger inhibition via internal attention.

In Experiment 3, we aimed to probe how internal attention to multiple features affects learning about novel objects. Attending to two features in VSTM produced even greater suppression of the unpracticed feature than attending to one feature twice. What might

account for this finding? We propose that memory retrieval functions as a form of reinforcement, and that competing representations that are repeatedly passed over for retrieval reinforcement are particularly susceptible to forgetting. Upon initial exposure to an object, there is no reason to privilege one feature dimension over another when forming a representation (absent systematic differences in salience or goals). Cued retrieval may disrupt this equal weighting by inducing competition among features vying for retrieval. Ultimately, the competition is resolved by the successful selection of the cued feature and the inhibition of uncued features. As a result, the retrieved feature enjoys a boost to its market share in long-term memory. Other dimensions that came to mind at the time of retrieval suffer a decrement, consistent with mechanisms of competition-based suppression that have been proposed to account for memory enhancement and suppression in other paradigms (Anderson et al., 1994, Anderson & Spellman, 2005, Norman & Newman, 2007). Experiment 1 is consistent with this account, since practiced features were remembered better than unpracticed features.

According to this framework, the representational strength of target and distractor features modulates the degree to which retrieval-based shifts in memory occur, such that unpracticed features that compete strongly with the practiced feature suffer greater impairment than weakly competing features. If the same feature is practiced for a second time, we expect unpracticed features to compete less fiercely during this second retrieval event because of their weakened status, and thus are not suppressed to the same extent.

If a different feature is probed upon the second exposure to an object, an accurate report of this feature may be more difficult because it was previously inhibited. Moreover, other previously unpracticed features may strongly compete during this second retrieval attempt,

as their strength may approximate that of the probed feature, in addition to the relatively strong previously probed feature. As a consequence of this additional round of competition, never probed features may experience even further marginalization. Consistent with this prediction, in Experiment 3 we observed worse test performance on unpracticed features of objects for which multiple features had been previously practiced, relative to those for which a single feature was practiced multiple times (Figure 1.7).

Overall enhancement and suppression could reflect two different types of changes in memory: retrieval practice of one feature could reduce the precision or the accessibility of the representations of other features. These two changes are not mutually exclusive, and their relative contribution can be estimated from continuous report data using a standard mixture modeling procedure (Zhang & Luck, 2008). Each response is treated: (1) as arising from a representation with a certain precision when accessible, or (2) as a random guess when inaccessible. Modeling the data this way generates one parameter estimate representing the probability that responses were guesses (*g*) and a second parameter representing the standard deviation of the responses that were not guesses around the correct value (*sd*). This procedure was applied to the data from all three experiments. The parameter that most reliably differed between tests of practiced vs. unpracticed features was *g*, the probability of guessing. In contrast, the estimated precision of feature memory was not reliably different across conditions. These findings suggest that retrieval principally affects whether the features of an object are accessible, not the precision with which they are stored.

So far, we have interpreted enhancement and suppression in VLTM to be a consequence of selective feature retrieval from VSTM. An alternative possibility is that observers learned which feature was relevant for a given object during its initial presentation in the practice

phase, and then strategically encoded this feature into VSTM on subsequent presentations, resulting in better VLTM performance in the test phase. According to this view, our findings may at least partly reflect differences in encoding rather than retrieval.

We consider this "selective encoding" account to be unlikely for the following reasons: First, observers were equally (Experiments 1 and 2) or more (Experiment 3) likely to be tested in VLTM on an unpracticed feature as they were on a practiced feature, reducing the incentive to strategically encode practiced features. Second, VSTM performance should have improved upon repeated presentations as observers learned over time which feature would be probed, but this improvement was not apparent after eliminating the possibility of motor learning in Experiments 2 and 3 (see Results). Third, VLTM performance should have been more related to VSTM performance on repeated (when selective encoding was possible) vs. initial presentations, but this difference did not emerge in any experiment (ps > 0.273). Fourth, even though observers were more likely to practice the same feature repeatedly in Experiment 3, VSTM performance was no worse when a new (Variable) vs. repeated (Constant/Switch) feature was probed in the second presentation (p > 0.197). Finally, insofar as the repeated practice of the same feature induced a selective encoding strategy, VSTM performance should have been better when the likelihood of repeated practice was higher in Experiments 1 and 2 (100%) vs. Experiment 3 (67%), but no such differences emerged for the second presentation (ps > 0.361). Taken together, all of these findings are incompatible with the alternative selecting encoding interpretation of the observed VLTM effects.

Considering the role of memory strategies raises an interesting connection to the literature on directed forgetting (MacLeod, 1998). For instance, after being presented with a VSTM array containing two objects, being post-cued to forget one of the objects leads to

better memory for the remaining object in the array (Williams et al., 2013). Likewise, being post-cued to remember a subset of objects boosts VSTM for these objects relative to receiving no cue (Williams & Woodman, 2012). The relationship between instructing observers to maintain a feature in VSTM ('directed remembering') vs. to retrieve and report a feature from VSTM ('selective retrieval') awaits further study. These two modes of internal selection may have qualitatively similar consequences for memory (suppression and enhancement), but these consequences could arise from the same or different mechanisms. On the one hand, directed forgetting engages metacognitive strategies to voluntarily bias long-term memory, in contrast to the present case, where the primary task was to immediately report a feature from VSTM. On the other hand, similar executive control processes invoked to resolve competition may play a similar role in both kinds of enhancement/suppression, regardless of whether they are engaged intentionally (Anderson, 2003). Indeed, we have interpreted our findings in terms of competition among features during selective retrieval from VSTM, but the directed forgetting literature suggests additional mechanisms that should also be considered (MacLeod, 1998, Sahakyan et al., 2009).

The consequences of selective internal attention to features might extend beyond subsequent feature memory. For example, to the extent that retrieval history directly alters the weighting of features within object representations, perceptual sensitivity to previously retrieved features may be heightened upon subsequent encounters with that object (Goldstone, 1998). Relatedly, tuning of object representations may result in previously retrieved features appearing more salient, such that the presence of these features guides the deployment of attention across a crowded display to facilitate localization and identification. Finally, the competition induced in the current study may reflect the random and indepen-

dent assignment of feature values to objects; introducing meaningful dependencies across dimensions may result in more generalized object-based benefits of feature retrieval. These are all consequences of selective feature retrieval for subsequent processing of the same object. Similar dynamics also operate at the level of objects themselves, with attention enhancing memory of selected objects and suppressing memory of inhibited objects (Fox, 1995b, Lepsien & Nobre, 2006, Nobre et al., 2004b, Tipper, 1985). An open question concerns how selective feature retrieval for one object biases processing of other objects, either when presented concurrently or encountered in the future.

1.6 Conclusions

Taken as a whole, the present results reveal the highly interactive nature of attention, learning, and memory, as well as how 'cognitive actions' can have long-lasting consequences for the organization of object representations. Although we focused on internal attention, cognitive operations that transform mental representations in the course of 'reading out' their contents appear to be pervasive. For example, when making a decision, although our preferences obviously guide our choices, our choices can also shift our preferences (Brehm, 1956, Festinger, 1957, Lieberman et al., 2001). Just as physical objects are neither static nor inert, objects of perception that comprise the internal visual world are amenable to cognitive manipulation. In this way, even the simple act of selection during the initial stages of learning has the potential to leave enduring traces in memory.

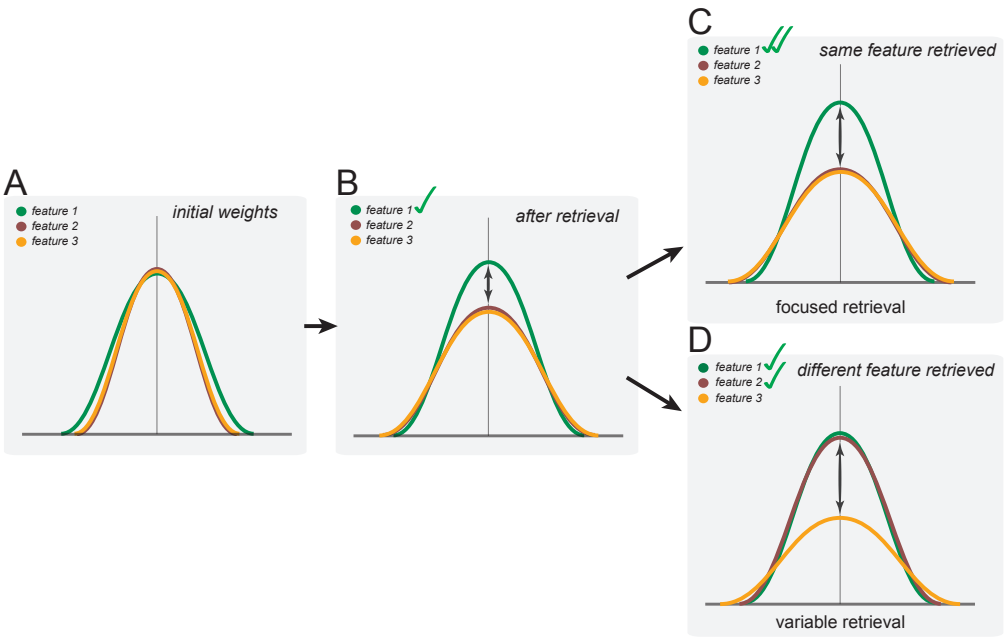


Figure 1.7: Interpretation. (A) Initially, attentional weights allocated to each feature dimension when forming a representation are roughly balanced. (B) Cued retrieval to a specific feature dimension may disrupt this equal weighting by inducing competition between dimensions for representation in short-term memory. This competition may be resolved via the selection of the object's value on the cued dimension and inhibition of values from uncued dimensions. Crucially, the consequences of this selective internal attention are not restricted to short-term memory, but extend to how various feature dimensions of an object are prioritized in long-term memory. (C) If the same feature is retrieved again from short-term memory in a subsequent encounter with that object, the competition among dimensions may be less fierce because this feature is already more strongly represented. As a result, the representation of the uncued dimensions may be further suppressed only slightly. (D) If a different feature is cued upon a subsequent encounter, accurate retrieval of this cued dimension from short-term memory may be more difficult due to the stronger status of the previously retrieved feature and prior inhibition of the newly cued feature. Moreover, given the weaker status of the cued feature, other previously uncued features may compete more, resulting in the further marginalization of still uncued feature dimensions in long-term memory.

2

How do visual memories affect what we pay attention to?

Much of the visual information we encounter is highly familiar. We inhabit the same places over time, such as our homes and offices, and these environments are populated by the same people and things. Because of this stability, memory for prior experiences can help us

achieve our behavioral goals, such as finding lost keys or a particular spice in the cupboard. This influence of memory on behavior has been construed as a form of top-down attention (Awh & Jonides, 2001), whereby actively maintaining information in working memory increases the priority of matching information in the world, even when task-irrelevant (Soto et al., 2008b). For example, when holding a colored shape in mind for a secondary delayed-match-to-sample task, the appearance of that shape as a distractor during a primary orientation-discrimination task impairs performance (Soto et al., 2007, Downing, 2000, Olivers et al., 2006, Carlisle et al., 2011).

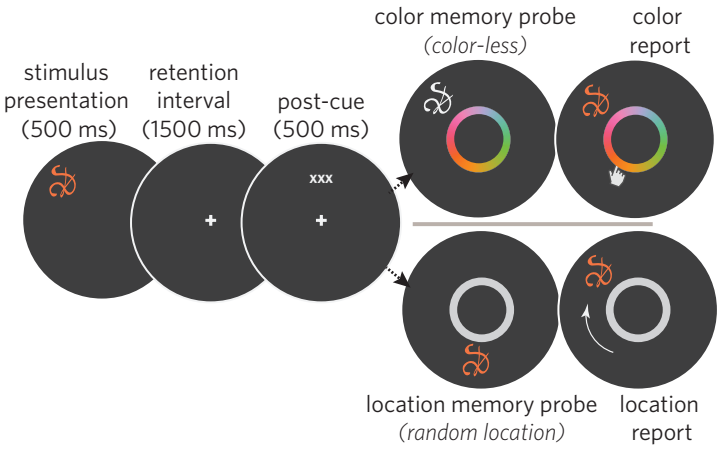


Figure 2.1: On each trial, one shape was presented in a unique color and location. After offset, a post-cue ("xxx" replaced by "color" or "location") prompted observers to report one of these two features by continuously adjusting a memory probe until it matched the original stimulus. Objects were viewed multiple times and the same feature was post-cued each time.

This prior work shows that recently experienced information, when actively maintained in working memory, gets prioritized if it appears again in the world. However, this mechanism for memory-guided attention has two limitations. First, there are often gaps in time longer than the span of working memory between repeated encounters with the same in-

formation (e.g., leaving home in the morning and returning in the evening). Second, when pursuing the goal of finding a target, other information associated with the target but not directly held in working memory might be useful (e.g., looking for a spice based on its label but also knowing its color and texture). In both cases, what influences attention is information from long-term memory retrieved coincidentally with the goal.

Here we characterize the influence of visual long-term memory (VLTM) on the guidance of attention. We show that information that was encoded in the past, but that is no longer present or relevant to the task, nevertheless influences visual search (Experiment 1) and captures attention (Experiment 2) by way of learned associations with the contents of working memory. This contrasts with previous studies showing biases in attention when information in long-term memory is directly presented again (Christie & Klein, 1995), or when explicitly searching for information in working memory is the goal of the task (Summerfield et al., 2006, Stokes et al., 2012). In addition, our study complements a growing literature on attentional guidance from long-term memory that has focused on how visual search is biased by pre-existing associations in semantic memory (Moores et al., 2003, Olivers, 2011, Rappaport et al., 2013). By training novel and arbitrary visual feature associations within an experimental session, our study seeks to link episodic memory to attention.

2.1 Experiment 1: Guidance in Visual Search

The goal of this experiment was to investigate how VLTM for features associated with a target influences visual search. We tested this by manipulating whether the target appeared with or without those features. We also examined whether such guidance results from automatic reactivation or explicit retrieval of associated features by manipulating their conscious

accessibility in VLTM.

2.1.1 Methods

Participants

Thirty naïve observers (18 women, mean age=22.1y) participated. In this and the next experiment, observers provided informed consent to a protocol approved by the Princeton Institutional Review Board, reported normal or corrected-to-normal visual acuity and color vision, and received course credit or $12 per hour as compensation.

Stimuli

Eight "alphabets" of eight shapes each were used (from Fan & Turk-Browne, 2013). To create novel objects, each shape was assigned a unique angular location (radial eccentricity=8°) and color (CIE L*a*b* space centered at L=54/a=18/b=-8; radius=59°). Stimuli were presented on a CRT monitor 70cm from the observer, controlled by MATLAB and the Psychophysics Toolbox (http://psychtoolbox.org).

Procedure

The experiment contained eight blocks, each employing one alphabet. Within each block, observers completed three phases: encoding, search, and recall. In the encoding phase, each object was presented three times in a randomized order, with the constraint that every object appeared once before any object repeated. On every trial (Figure 2.1), an object was presented for 500ms, followed by a 1500ms retention interval. Observers were then post-cued to report the object's color or location. We previously showed that this form of retrieval prac-

tice robustly affects the conscious accessibility of features in VLTM, both enhancing the accessibility of practiced features and reducing the accessibility of unpracticed features (Fan & Turk-Browne, 2013). Observers continuously adjusted a memory probe until it matched the original object on the post-cued feature; the shape and uncued feature were presented veridically (e.g., if cued to report the color of a shape, it was shown in the correct location and vice versa). Each object was assigned to either the color-practice or location-practice condition; these assignments were randomized over objects and observers. Because the influence of VLTM on search was of primary interest, only objects whose features were successfully reported during the encoding phase were retained for analysis. This was defined as report error within 1.5x the interquartile range beyond the 1st and 3rd quartiles (Fan & Turk-Browne, 2013a, Frigge et al., 1989). On average, 3.43 of the 64 objects were excluded per observer. This rate did not differ between color-practice and location-practice items ($t<1$).

In the search phase, observers searched for a cued target shape in cluttered arrays of colorful distractors generated from fragments of novel shapes. The cue consisted of a monochromatic, centrally presented version of the shape. Each shape was cued twice, once present and once absent. The assignment of objects to target conditions was counterbalanced (Figure 2.2A). This resulted in four conditions for present trials (location match/color match): target matched original object's color and location (+/+); target matched original object's location but appeared in a novel color (+/–); target matched original object's color but appeared in a novel location (–/+); target appeared in a novel color and location (–/–). Novel features were randomly selected > 90° from the original value in feature space. Observers were instructed to determine as quickly and accurately as possible whether the shape was present or absent — i.e., regardless of its color or location. The array remained onscreen until response.

We tested two hypotheses: First, if VLTM for associated features guides search for the target shape, response times (RTs) should be faster when there is a location and/or color match. Second, if these features are reactivated automatically, RT effects should not only be present when a matching feature was practiced, and thus more consciously accessible (Fan & Turk-Browne, 2013), but also when a matching feature was not practiced (e.g., location match for color-practiced objects). In the recall phase, observers were prompted to report the practiced feature of each shape, as verification that VLTM had been formed.

2.1.2 Results and Discussion

Encoding phase

Accuracy exceeded chance (root mean squared error [RMSE] = 103.9°) for all three presentations of a shape for both color (RMSEs < 12.7°; $t(29)$s > 139.2, ps ≪ 0.001) and location (RMSEs < 3.61°; $t(29)$s > 711.8, ps ≪ 0.001). We analyzed errors using a 2 (dimension: color, location) x 3 (presentation number: first, second, third) repeated-measures ANOVA (Table 2.1). There was a main effect of feature dimension ($F(1,29) = 269.0$, p ≪ 0.001, $\eta 2 = 0.903$), with higher accuracy for location than color. There was no main effect of presentation number ($F(2,58) = 0.677$, $p = 0.512$, $\eta 2 = 0.023$) or interaction ($F(2,58) = 0.769$, $p = 0.468$, $\eta 2 = 0.023$), suggesting that repeated practice did not boost the fidelity of observers' already accurate reports.

Search phase

The primary dependent measure of interest was RT. As in other visual search tasks, RTs were slower for target-absent trials (M=3.33s, SD=1.88s) than target-present trials (M=1.51s,

Table 2.1: Encoding phase accuracy. Mean RMSE in degrees (and standard error of the mean [SEM]) for the report task, by practiced dimension and presentation number.

		Presentation Number		
		1	2	3
s Practiced	Color	12.7° (0.59°)	12.3° (0.51°)	12.2° (0.66°)
Dimension	Location	3.21° (0.14°)	2.76° (0.14°)	3.61° (0.14°)

SD=0.35s; $t(29) = 6.03$, $p < 0.001$, $d = 1.10$). Because conditions were defined by the color and location of the target, only correct target-present trials were included in subsequent RT analyses. RTs greater than 3 SDs from the mean in each condition or longer than 10s were trimmed (0.26% of trials). To mitigate the influence of extreme values, single-trial RTs were also log-transformed prior to computation of the mean, and the mean was then transformed back using the exponential function. None of these steps affected the overall pattern of results.

To test our first hypothesis that feature associations in VLTM guide search, we analyzed RTs using a 2 (location match) x 2 (color match) repeated-measures ANOVA (Figure 2.2B). There was a main effect of location match ($F(1,29) = 12.8$, $p = 0.001$, $\eta 2 = 0.306$), but no main effect of color ($F(1,29) = 1.31$, $p = 0.262$, $\eta 2 = 0.043$). This suggests that VLTM specifically guided spatial attention to the search target, perhaps reflecting the spatial demands of the visual search task or the more precise encoding and better recall of location associations. However, a reliable interaction ($F(1,29) = 4.19$, $p < 0.05$, $\eta 2 = 0.126$), combined with a significant ordinal interaction ($t(29) = 3.69$, $p < 0.001$, $d = 0.674$; see (Strube & Bobko, 1989), revealed a super-additive benefit of matching colors for the location effect. This provides tentative evidence that color associations in VLTM may have partially guided attention, in

conjunction with location associations.

To test our second hypothesis that feature associations in VLTM can guide visual search automatically, we compared the location match effect in RT between location- and color-practice items, finding no difference as a function of practice ($F(1,29) = 0.216$, $p = 0.646$, $\eta 2 = 0.007$). Indeed, the location match effect was reliable for color-practice items when analyzed separately ($F(1,29) = 13.0$, $p = 0.001$, $\eta 2 = 0.310$), suggesting that conscious accessibility of features in VLTM did not modulate the degree to which VLTM influences spatial attention during visual search.

In order to examine how location and color associations influence search, the target shape appeared in its original location half of the time and in its original color half of the time. Therefore, although observers were instructed to search for the shape itself, it could have been beneficial to strategically retrieve and search for these features. To test this possibility, we examined the prevalence of this strategy and its effect on the results. During debriefing, 6 out of 30 observers (20%) reported using a strategy of retrieving and searching for associated locations and/or colors. When these observers were excluded from analysis, however, we obtained the same pattern of results, including: the main effect of location ($F(1,23) = 12.2$, $p = 0.002$, $\eta 2 = 0.347$), the significant ordinal interaction with color ($t(23) = 3.19$, $p = 0.004$, $d = 0.652$), and no modulation of the location effect by practiced feature ($F(1,23) = 0.001$, $p = 0.979$, $\eta 2 < 0.001$). These results lend further support to the idea that attentional guidance by VLTM can occur automatically.

For completeness, we also analyzed the accuracy data from the search phase (Table 2.2). We quantified accuracy with A', a nonparametric measure of sensitivity (Grier, 1971). A 2 (location match) x 2 (color match) repeated-measures ANOVA revealed no main effects of

location match ($F(1,29) = 0.644$, $p = 0.423$, $\eta 2 = 0.022$) or color match ($F(1,29) = 0.687$, $p = 0.414$, $\eta 2 = 0.023$), and no interaction ($F(1,29) = 0.878$, $p = 0.356$, $\eta 2 = 0.029$). These results argue against the possibility that the RT effects reflected a speed-accuracy tradeoff.

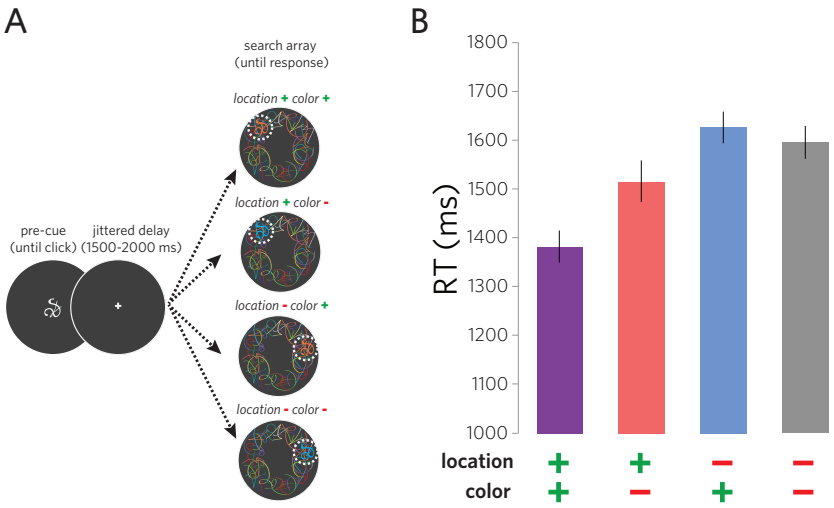


Figure 2.2: Search phase of Experiment 1. Observers determined whether a cued shape was present or absent within a cluttered display, regardless of its color or location (bottom). Mean log-transformed search RTs for correctly detected targets, as a function of whether the target matched the color and/or location of the cued shape from encoding (top). Error bars reflect ±1 S.E.M.

Recall phase

During the recall phase, accuracy exceeded chance for both color (RMSE = 61.4°; $t(29) = 11.4$, $p \ll 0.001$, $d = 2.08$) and location (45.5°; $t(29) = 21.0$, $p \ll 0.001$, $d = 3.84$), with better memory for location than color ($t(29) = 5.82$, $p < 0.001$, $d = 1.06$). These results verify that observers encoded the shapes and their practiced features into VLTM.

Table 2.2: Search phase accuracy. Mean A', hit rate, and false alarm rate by condition (with SEM).

Location	Color	Accuracy		
		A-prime	Hit rate (%)	False alarm rate (%)
+	+	0.956 (0.010)	88.3 (2.24)	4.0 (1.60)
+	-	0.943 (0.019)	86.7(2.84)	4.8 (2.48)
-	+	0.944 (0.019)	87.1 (2.43)	4.4 (2.34)
-	-	0.946 (0.016)	86.3 (2.19)	4.2 (1.73)

2.2 Experiment 2: Attentional Capture from VLTM

Experiment 1 suggested that VLTM for the location of an object is reactivated during visual search for the object. Experiment 2 sought to expand these findings in three ways. First, we wanted to better assess VLTM guidance for color. In Experiment 1, the presence of a matching color from VLTM only influenced visual search insofar as the location also matched VLTM. Here we use a simpler paradigm that has previously revealed attentional capture for colors and shapes in working memory (Soto, Humphreys, & Rotshtein, 2007).

Second, we aimed to further test the automaticity of such guidance by making VLTM associations entirely task-irrelevant. In Experiment 1, associated locations and colors could be construed as task-relevant because they were each useful on half of trials for finding the target. Observers were not informed about this, and debriefing data suggested that strategic use of these relationships was uncommon, as well as unnecessary to observe guidance. Nevertheless, to more definitively rule out such a strategy, we employ a dual-task paradigm in which associated colors serve as task-irrelevant distractors.

Third, we sought to verify that our practice manipulation affected the conscious accessi-

bility of features in VLTM. In Experiment 1, the location guidance effect was observed for color-practice objects, which we interpreted as support for automaticity because the unpracticed locations underlying this guidance were assumed to be less consciously accessible. This assumption was based on our prior work (Fan & Turk-Browne, 2013), and we were unable to independently verify it because the final recall phase only tested practiced features. Here we test recall of both practiced and unpracticed features in a subset of observers to replicate this prior work and support the claim that unpracticed features were less accessible.

2.2.1 Methods

Participants

Sixty naïve observers (34 women, mean age 19.2y) participated.

Procedure

For each observer, six alphabets each containing ten shapes were used, for a total of six blocks. Within each block, a similar three-phase design was used: encoding, dual-task, and recall. There were three differences in this experiment: First, each object was presented five vs. three times during encoding, which we thought might benefit learning of feature associations, without extending the duration of the experimental session beyond one hour. Each object was assigned to either location- or color-practice. Again, only objects whose practiced features were successfully reported during encoding were retained for analysis (2.9 of the 60 objects excluded per observer). Second, the search task was replaced by a dual-task consisting of an outer-loop delayed-match-to-sample task and an inner-loop orientation-discrimination task (Figure 2.3A). The delayed-match-to-sample task required that a monochro-

matic, centrally presented version of a shape be maintained in working memory (cue = 2000ms, delay = 700ms+orientation discrimination RT; modeled on (Soto & Humphreys, 2007)) and compared to a test shape (50% match). During the delay, two horizontally offset Gabor patches were presented (frequency = 2cpd; eccentricity = 6.8°) for the orientation-discrimination task. One (target) patch was randomly tilted 26° (clockwise/counterclockwise) from vertical and the other (distractor) patch was vertical. The location of the target was randomized, and observers judged its tilt direction. Additionally, one of the patches appeared in the color associated with the delayed-match-to-sample target in VLTM and the other in a different color (>90° away). Each shape served as the delayed-match-to-sample target twice: once where the target patch in the orientation-discrimination task matched the associated color (valid), the other where the distractor patch matched the associated color (invalid). Thus, the associated color predicted neither the location nor orientation of the target and was therefore completely task-irrelevant. Before the experiment, observers were trained on both tasks to the same performance criterion: 90% correct with RTs < 1000ms ((Soto et al., 2007)). This phase allowed us to test a key hypothesis — that VLTM for the associated color is reactivated during the delayed-match-to-sample task, drawing attention to the matching patch in the orientation-discrimination task, and speeding RTs for valid vs. invalid trials. Such a result would license strong claims about the automaticity of guidance: not only was the color association irrelevant to the delayed-match-to-sample task, the shape cue was irrelevant to the orientation-discrimination task. Third, to verify that our manipulation of conscious accessibility of features in VLTM was successful, we ran two cohorts of observers in this experiment (each with N = 30, equal in size to Experiment 1). The first cohort was tested on either the practiced or unpracticed feature of each object (counterbal-

anced across objects) during the recall phase. Comparing recall performance as a function of practice allowed us to verify that unpracticed features were less consciously accessible. The second cohort was always tested on the practiced feature of each object in the recall phase to mirror Experiment 1. The two cohorts only differed in the recall phase (after the critical dual-task phase) and thus we did not expect this between-subjects manipulation to influence the critical dual-task results. Therefore, we pooled observers from both cohorts into the final sample, though we also verified with an ANOVA that the amount of guidance did not differ between groups. To compare recall performance for practiced and unpracticed features in the first cohort, we used an approach from Fan & Turk-Browne (2013a), who found that unpracticed features were less accessible (the basis for our assumption in Experiment 1). Specifically, we fit a model to the distribution of errors that observers made when reporting one or the other feature in the recall phase (Zhang & Luck, 2008, Wilken & Ma, 2004). Maximum likelihood estimation was used to partition errors into two component distributions (Equation 2.1): (1) a uniform distribution reflecting the probability of guessing (i.e., 1 – probability of successful retrieval from memory) and (2) a von Mises distribution (an approximation to a wrapped normal distribution) centered on the actual feature value whose variance provides a measure of the precision of memory for the feature.

$$p(\hat{\theta}) = \beta\phi_{\sigma}(\hat{\theta} - \theta) + (1 - \beta)\frac{1}{360} \tag{2.1}$$

Here, β represents the probability of successful retrieval from memory, and $\phi\sigma$ represents the von Mises distribution centered on the actual feature value θ with standard deviation σ. Because of the small number of recall phase trials collected per test condition per observer (30 trials each), errors were pooled across observers within each condition for subsequent

Table 2.3: Encoding phase accuracy. Mean RMSE (and SEM) in the report task by practiced dimension and presentation number.

		Practiced Dimension				
		1	2	3	4	5
Presentation	Color	15.5° (1.15°)	15.6° (0.83°)	15.1° (0.96°)	15.2° (1.03°)	14.4° (0.80°)
Number	Location	10.2° (0.40°)	10.8° (0.51°)	10.6° (0.55°)	10.4° (0.51°)	10.2° (0.46°)

analysis, providing sufficient statistical power to derive robust model estimates. Random-effects reliability was established across observers by iteratively resampling observers with replacement 2500 times to derive 95% confidence intervals for each model parameter in each condition (Efron & Tibshirani, 1986).

2.2.2 Results and Discussion

Encoding phase

During the encoding phase, performance was better than chance at all five item presentations for both color (RMSEs < 15.7°; all $t(59)$s > 77, ps ≪ 0.001) and location-practiced items (RMSEs < 10.8°; all $t(59)$s > 169, ps ≪ 0.001). We analyzed encoding phase errors using a 2 (dimension: color, location) x 5 (presentation number) repeated-measures ANOVA (Table 2.3). Accuracy was again higher for the location vs. color dimension ($F(1,59) = 40.2$, $p<0.001$, $\eta 2 = 0.405$), and there was no effect of presentation number ($F(4,236) = 1.06$, $p = 0.378$, $\eta 2 = 0.018$).

Dual-task phase

Only trials where responses to both the orientation discrimination and delayed-match-to-sample tasks were correct (93.3%) and less than 1000ms (93.9%) were included in RT analyses (from Soto et al., 2007). The delayed-match-to-sample task served to induce maintenance of a shape with color associations, so beyond showing that participants were highly accurate in this task overall (Table 2.4; vs. chance of 50%, $t(59) = 87.2$, $p \ll 0.001$, $d = 11.3$), it is not analyzed further and we focus analysis on the orientation-discrimination task.

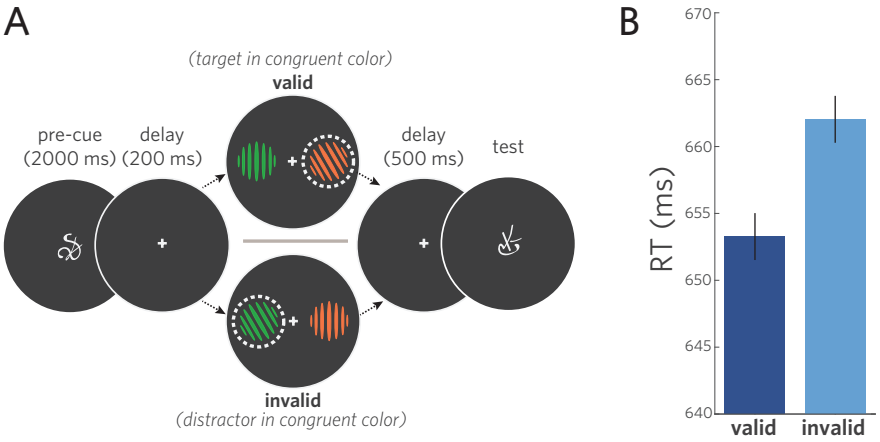


Figure 2.3: Dual-task phase of Experiment 2. (A) While maintaining a (monochromatic) shape in working memory, observers discriminated the orientation (clockwise/counterclockwise) of the tilted Gabor (target). The color of this patch was associated with the shape on half of trials (valid). On the remaining trials, the vertical Gabor (distractor) appeared in this color (invalid). (B) Mean orientation-discrimination RTs as a function of validity. Error bars reflect ±1 S.E.M.

Consistent with our hypothesis, orientation discrimination was slower for invalid (662ms) vs. valid trials (653ms; $t(59) = 2.57$, $p = 0.023$, $d = 0.332$), suggesting that the color associated with the shape in working memory captured attention (Figure 2.3B). The color validity effect did not differ between location- and color-practice items ($F(1,59) = 1.10$, $p = 0.298$, $\eta 2 = 0.018$), similar to what was found in Experiment 1. The two observer cohorts did not dif-

Table 2.4: Dual-task phase accuracy. Orientation-discrimination and delayed-match-to-sample (DMS) task performance (% correct) by condition.

			Accuracy	
			Orientation Discrim.. (%)	DMS(%)
	Color-Prac.	Valid	96.6 (0.67)	95.9 (0.66)
		Invalid	97.0 (0.54)	96.1 (0.57)
Cond.	Location-Prac.	Valid	97.1 (0.48)	96.4 (0.62)
		Invalid	96.5 (0.58)	96.3 (0.75)

fer in overall RT ($F(1,58) = 0.061$, $p = 0.805$, $\eta 2 = 0.001$), nor did the color validity effect differ across cohorts ($F(1,58) = 0.029$, $p = 0.864$, $\eta 2 < 0.001$). Moreover, accuracy did not differ between valid and invalid trials ($t(59) = 0.051$, $p = 0.959$, $d = 0.007$), inconsistent with a speed-accuracy tradeoff (Table 2.4).

We interpret these findings as reflecting attentional capture from the memory-matching color on invalid trials (consistent with Soto, Humphreys, & Rotshtein, 2007). However, another possibility is that the memory-matching color facilitated attention to the target on valid trials. Future experiments could disentangle these processes with a neutral condition (i.e., search arrays without the memory-matching color). Nevertheless, the key goal of this experiment was to show that, beyond the salience of the current display and the identity of cues being maintained in working memory, information from VLTM can also influence attention. The RT difference between valid and invalid trials, whether driven by capture or facilitation, helps achieve that goal.

Table 2.5: Recall phase performance. Model parameter estimates for responses when tested on the practiced feature vs. unpracticed feature.

	Probability of Retrieval		Precision of Memory	
	MLE	95% CI	MLE	95% CI
Practiced Feature	0.699	[0.613, 0.779]	14.3°	[12.6°, 16.0°]
Unpracticed Feature	0.626	[0.540, 0.714]	15.7°	[14.0°, 17.5°]

Recall phase

During the recall phase, both color (RMSE = 51.9°; $t(59) = 20.3$, ps $\ll 0.001$) and location (RMSE = 51.5°; $t(59) = 18.4$, ps $\ll 0.001$) accuracy exceeded chance performance, and they were not different from one another ($t(59) = 0.167$, $p = 0.867$, $d = 0.022$), demonstrating successful encoding of practiced features into VLTM.

Confirming that our practice manipulation influenced recall (without affecting guidance), the probability of retrieval (β) was lower for the unpracticed feature than for the practiced feature in the cohort who was tested on both ($p = 0.003$; Table 2.5). There was no such difference in precision (inversely related to σ) however ($p = 0.223$), suggesting that practice affected the conscious accessibility of memories but not their fidelity.

2.3 General Discussion

This study aimed to elucidate some ways in which long-term memory influences visual selection. To accomplish this, we associated multiple features with novel shapes in VLTM, and then tested how these associations subsequently biased attention when this information was no longer task-relevant. Experiment 1 showed that VLTM for the associated location of

a target guided spatial attention during visual search for the target, even when this location was task-irrelevant. Experiment 2 extended these findings by showing that VLTM for the associated color of a target induced attentional capture in a different task.

These results build on previous work that has examined the relationship between memory and attention (Awh, Belopolsky, and Theeuwes, 2012; Hutchinson and Turk-Browne, 2012). In studies of contextual cueing, repeated pairings of a spatial context with a target location can speed visual search (Chun & Jiang, 1998, Endo & Takeda, 2004). However, in contrast to such studies — in which associations are encoded and tested in the same task — Experiment 1 demonstrated guidance in visual search after mere exposure to objects in particular locations, despite the fact that locations were task-irrelevant and variable. There are many other compelling examples of attentional guidance by task-irrelevant information in memory (Soto et al., 2007, Downing, 2000, Olivers et al., 2006, Carlisle et al., 2011), but these typically depend on working memory for recently viewed features rather than automatic reactivation of these features from VLTM.

The present findings suggest a common mechanism for these diverse examples of memory-guided attention, by which automatic retrieval of an object's features from long-term memory into working memory biases selection toward items in the world that at least partially match reactivated feature representations. This interpretation is consistent with modal models of memory (Shiffrin & Atkinson, 1969a), in which working memory serves as the critical juncture between long-term representations and ongoing sensory processing. Moreover, it makes novel predictions about the scope of VLTM guidance of attention, insofar as capacity limitations in working memory define a bottleneck for reactivation and guidance.

Interestingly, our manipulation of the conscious accessibility of features in long-term

memory did not modulate the degree of attentional guidance. This finding suggests that while selective attention during encoding constrains subsequent explicit recall of object features (Fan & Turk-Browne, 2013b), unattended features may be encoded at some level such that they can still influence behavior. In particular, an initial lack of attention may impair later controlled retrieval of unattended features from long-term memory, but these features may nevertheless be automatically reactivated based on associations with information in working memory. Ultimately, such automatic retrieval mechanisms may be crucial for efficiently recruiting long-term knowledge in the pursuit of ongoing perceptual goals.

3

How do visual memories affect our perceptual judgments?

What we see is intuitively thought to be determined by the light reaching our eyes, but it may also be influenced by memories for what we have seen before. This supplemental role of memory in perception might be especially needed under impoverished viewing condi-

tions (e.g., dim lighting). For instance, it may be easier to recognize the face of a close friend in a dark movie theater than that of an acquaintance, even though we may have no problem identifying either under normal illumination.

There are multiple ways in which past experience can influence present perceptual performance. Benefits may derive from extensive practice on specific perceptual tasks, as in the case of perceptual learning (Sagi, 2011). For example, consistent exposure to dots moving in one direction leads to enhanced motion perception in that direction (Ball & Sekuler, 1987, Watanabe et al., 2001). Moreover, over a lifetime of experience, real-world objects possessing a 'canonical' color (e.g., bananas) are perceived to have that color even when presented achromatically (Hansen et al., 2006).

Benefits may also accrue from minimal experience, such as a single exposure to a word or object, which results in faster and more accurate identification upon repetition (Tulving & Schacter, 1990), even lasting for weeks (Cave, 1997). One remarkable feature of such priming is that it does not require awareness that an item has been seen before. When encountering a repeated stimulus, people may erroneously attribute their perceptual fluency to aspects of current viewing conditions (e.g., presentation duration, the level of noise) rather than to memory (Witherspoon & Allan, 1985, Jacoby et al., 1988). Furthermore, amnesic patients, who are not able to consciously recall recent experiences, nevertheless can exhibit priming (Warrington & Weiskrantz, 1974).

Taken together, this prior work shows that information encoded into long-term memory can automatically influence perception, partly by enhancing perceptual sensitivity for that information when it is presented again. Although this suggests that long-term memory interacts with current sensory inputs in some way to determine perception, how these

interactions occur is less well understood.

Recently, the field of visual memory has developed quantitative modeling techniques to characterize the strength and fidelity of visual representations (Zhang & Luck, 2008, van den Berg et al., 2012, Fougnie et al., 2012, Suchow et al., 2013). These tools have spurred advances in our understanding of diverse phenomena (Brady et al., 2011), including feature binding in visual working memory (Fougnie et al., 2010, Bays et al., 2011), encoding precision (Bays & Husain, 2008, van den Berg et al., 2012), false memory and swapping effects (Bays et al., 2009), visual long-term memory (VTLM) encoding and retrieval (Fan & Turk-Browne, 2013a, Brady et al., 2013, Williams et al., 2012), attentional orienting (Golomb et al., 2014), and ensemble perception (Brady & Alvarez, 2011).

Here we use these tools to examine how VLTM representations and current sensory evidence jointly determine perceptual judgments. As a case study of when these information sources may be most likely to interact, we asked observers to judge the appearance of a familiar object near the threshold of awareness. Specifically, we first associated shapes with unique colors in VLTM then later presented the shapes again, but now in unrelated colors. Observers estimated the color in which each shape appeared, using continuous report. We manipulated the amount of sensory evidence available during these later presentations by reducing stimulus duration (from 150ms to 10ms). We hypothesized that reducing the availability of sensory information would increase reliance on memory, even though long-term memory was entirely task-irrelevant (as original and current colors were uncorrelated). By tightly controlling the similarity between the contents of VLTM and sensory experience (i.e., distance in color space) and the amount of sensory evidence (i.e., duration), this study provides novel quantitative insight into the influence of long-term memory on perception.

3.1 Methods

3.1.1 Participants

Across three experiments, 72 observers (49 women, age: 18-33y) who reported normal or corrected-to-normal visual acuity and color vision participated. The sample size was the same for each experiment (24 observers) and was determined based on piloting and previous research. Each observer received course credit or $12/hour as compensation and provided informed consent to a protocol designed in accordance with the principles expressed in the Declaration of Helsinki, and approved by the Institutional Review Board of Princeton University.

3.1.2 Stimuli

Eight 'alphabets' of twelve shapes each were used (Fan & Turk-Browne, 2013), for a total of 96 novel items. Each shape repeatedly appeared in the same randomlyT chosen angular location (radial eccentricity=8°) and color (CIE L*a*b* space centered at L=54/a=18/b=-8; radius=59°). Later, each shape re-appeared in the same location but in a new, independently sampled color. Stimuli were presented using a CRT monitor 70cm from the observer with MATLAB and PsychToolbox (http://psychtoolbox.org).

3.1.3 Procedure

The experiment contained eight blocks, each employing a unique alphabet. Blocks were subdivided into initial exposure and final test phases (Fig. 1). For each initial exposure trial, a shape was presented for 500ms, followed by a 1500-ms blank interval. An achromatic version

of the shape (the 'probe') then reappeared at the same location, along with a color response wheel (radius=4°). Observers were instructed to recreate the color in which the shape had just appeared by adjusting the probe's color using the color wheel, clicking to report the final selection (Wilken & Ma, 2004). The full alphabet of shapes was repeated 3x during initial exposure; the order of shapes within each repetition randomized. The exposure phase was identical across all three experiments. There were 36 trials/block x 8 blocks = 288 exposure trials in total. The final test trials were similar except for three changes: (1) Each shape was presented in a new, randomly chosen color. (2) Shapes were presented between forward and backward color-noise masks of jittered duration (150-225ms). Noise masks were trial-unique, ring-shaped arrays of randomly colored shape fragments that covered all possible angular locations (inner radius=6.5°; outer radius=9.5°). (3) The shape presentation duration was shortened to 150ms (Experiment 1), 50ms (Experiment 2), or 10ms (Experiment 3). Finally, a random subset of "invisible" trials in each experiment did not contain a shape at all (i.e., 0ms). Observers were always instructed to report the color of the shape as it appeared on the current trial, identical to the exposure phase, and were only informed about "invisible" trials during debriefing. There were 12 test trials/block x 8 blocks = 96 test trials in total.

Experiment 1

Twenty-four naïve observers participated (17 women, mean age=19.3y). Shapes were assigned in equal proportion to one of two visibility conditions, '150-ms' and '0-ms', and this assignment was counterbalanced between observers (48 shapes/condition).

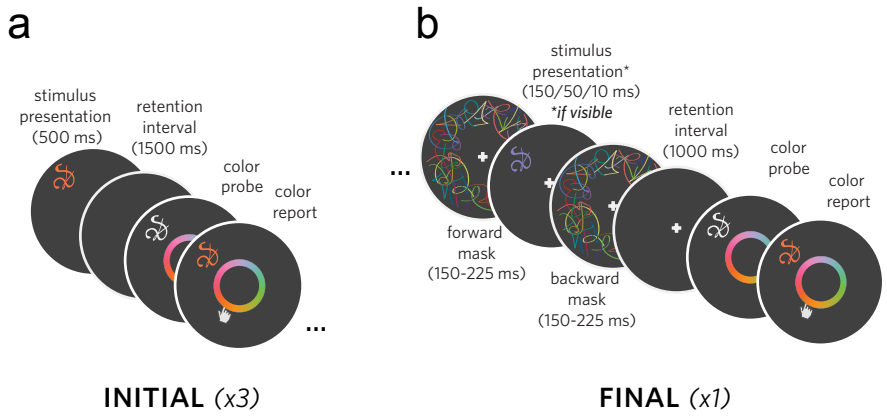


Figure 3.1: Task Display. (a) During initial exposure trials, a colored shape was briefly presented and then an achromatic version of the shape reappeared to probe observers for the color. (b) Final test phase trials were identical, other than the use of forward and backward masks, the duration of the initial shape, and the fact that the shape had previously appeared in a different color.

Experiment 2

Twenty-four naïve observers participated (15 women, mean age=21.1y). One-third of all items (32 shapes) were assigned to a '50-ms' condition and two-thirds (64 shapes) were assigned to a '0-ms' condition, to ensure there was sufficient statistical power in the '0-ms' condition to derive robust model parameter estimates. Upon reporting the color on each trial, observers indicated whether they detected the shape using a 4-point scale in response to the question "How confident are you that you saw the shape?": sure present, maybe present, maybe absent, or sure absent.

Experiment 3

Twenty-four naïve observers participated (17 women, mean age=22.5y). In this experiment, two-thirds of items (64 shapes) were assigned to a '10-ms' condition, and the remaining

one-third (32 shapes) to a '0-ms' condition, to ensure sufficient statistical power in the '10-ms' condition to derive robust model parameter estimates. As in Experiment 2, 4-point present/absent responses were collected.

3.1.4 Data Analysis

Due to the small number of final test trials collected per condition per observer, errors were pooled across observers within each condition for subsequent analysis, providing sufficient statistical power to derive robust model estimates. Random-effects reliability was established across observers using bootstrapping (see below).

Modeling performance on initial exposure trials

During initial exposure, there was only one feature associated with each shape. To model performance on these trials, we employed a mixture-modeling procedure that uses maximum-likelihood estimation to partition errors into two component distributions (Equation 1): a uniform distribution reflecting the probability of guessing (inversely related to the proportion of trials on which the response was "anchored" around the true color value); and a von Mises distribution of the anchored responses whose variance provides a measure of the precision (1/SD) of the color representation (Zhang & Luck, 2008).

$$p(\hat{\theta}) = \beta\phi_{\sigma}(\hat{\theta} - \theta) + (1 - \beta)\frac{1}{360} \tag{3.1}$$

β represents the frequency of responding from the color representation, where ϕ denotes the von Mises distribution centered on the correct feature value θ, with standard deviation σ.

Modeling performance on final test trials

On test trials containing a stimulus ('visible' trials), observers may have used one of two distinct sources of information to generate responses: long-term memory for the color associated with the shape on previous encounters or recent sensory information about the color of the shape on the current trial (Fig. 2). To estimate the likelihood of responding based on long-term memory, we adapted the mixture model above to include three components (Equation 2): a von Mises distribution for the 'original' color, describing the precision and probability of anchoring upon the long-term memory representation of feature value θ_{orig} from exposure; a von Mises distribution for the 'current' color, describing the precision and probability of anchoring upon the sensory representation of feature value θcurr that just appeared; and a uniform distribution describing the probability of guessing.

$$p(\hat{\theta}) = \beta_{orig}\phi_{\sigma_{orig}}(\hat{\theta} - \theta_{orig}) + \beta_{curr}\phi_{\sigma_{curr}}(\hat{\theta} - \theta_{curr}) + (1 - \beta_{orig} - \beta_{curr})\frac{1}{360} \qquad (3.2)$$

According to this model, perceptual judgment entails sampling directly from either long-term memory or sensory representations, resulting in a mixture of response distributions that reflect the independent origins of the underlying signals. This modeling approach has also been used to capture "swap" errors in continuous-report tasks, in which multiple items were concurrently presented but only a single target item was cued for report (Bays, Catalao, & Husain, 2009). In that case, swaps occurred when a distractor item from the current trial was reported, as opposed to our case, where swaps reflect a substitution from long-term memory.

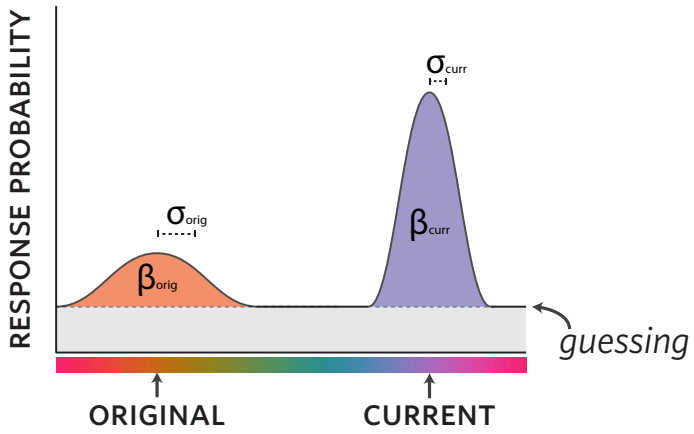


Figure 3.2: Modeling Approach. Responses across observers in the final test phase were modeled as a mixture of three component distributions: a uniform distribution to capture guess responses and von Mises distributions to capture responses centered on the 'original' color of the shape from exposure and its 'current' color, respectively. □ = probability of anchoring on a color; □-1 = precision of anchored responses.

Assessing reliability of parameter estimates

As responses were pooled across observers, the above modeling approaches produced single sets of parameter estimates for each condition. In order to determine the random-effects reliability of these estimates within-experiment, we used a bootstrapping procedure (Efron & Tibshirani, 1986). On each of 1000 iterations, 24 observers' worth of data were randomly sampled with replacement from the original 24 observers, and the models were re-fit to this bootstrapped pooled dataset. The logic underlying this resampling approach is that insofar as the effect is reliable across observers, similar results will be obtained regardless of the subset of observers sampled on a given iteration. Across iterations, this procedure yielded an empirical distribution for each model parameter. These distributions were used to construct confidence intervals, as well as to assess model performance. We have used this approach in other published studies that involve single-trial analysis (Kim, Lewis-Peacock, Norman, &

Turk-Browne, 2014; Hindy & Turk-Browne, 2015).

Assessing the fit of the three-component model

To evaluate the general validity of the three-component mixture model (Equation 2), we compared its Akaike Information Component scores (AIC; (Akaike, 1974, Burnham & Anderson, 2004)) with those of two reduced versions of the model and a null model: (i) a two-parameter version containing only the current-color component (thus ignoring long-term memory), plus guessing; (ii) a two-parameter version containing only the original-color component (thus ignoring immediate sensory information), plus guessing; (iii) a zero-parameter null model, containing only the uniform-guessing component.

To assess variability in our group-level estimate of AIC, we used the same bootstrapping procedure as above to construct confidence intervals for the AIC scores of the full, reduced, and null versions of the model in each experiment (Figure 3.6). The p-value reported in the results as evidence that observers' performance exceeded that expected for random responding was defined as the number of iterations on which the null model attained a lower AIC (i.e., a better fit) than the full model, and is labeled pAIC. For invisible trials, the full model contained only the original-color component, plus guessing.

Comparing parameters across experiments

To evaluate the consequences of reducing sensory evidence on final test trials, we performed a randomization test across groups that assessed the statistical reliability of trends in model parameters across experiments. Such trends were quantified as the slope of the line fit to observed parameter values vs. stimulus duration (10, 50, 150ms). A null distribution of slope

Table 3.1: Model parameter estimates for responses during the initial exposure phase.

Experiment	Anchoring Probability		Precision	
	MLE	*95% CI*	*MLE*	*95% CI*
E1 (150ms)	0.962	[0.940, 0.981]	0.074	[0.067, 0.081]
E2 (50ms)	0.967	[0.955, 0.978]	0.071	[0.064, 0.078]
E3 (10ms)	0.974	[0.964, 0.983]	0.078	[0.072, 0.085]

values was derived by fitting a line 1000x to three random, equally sized subdivisions of the set of all observers. Two-tailed p-values were derived by evaluating the proportion of values in this distribution that exceeded the observed slope in magnitude. Follow-up analyses of pairwise differences in model parameters between experiments were conducted using a similar procedure. For a given pair of experiments, the null distribution of differences in parameter values was derived by fitting the model 1000x to two random, equally sized subdivisions of observers from both groups. Two-tailed p-values were derived by looking up the observed difference in this null distribution and identifying the proportion of iterations that exceeded it in magnitude.

3.2 Results

3.2.1 Initial Exposure Trials

We first assessed performance on initial exposure trials, to verify successful encoding of color associations into VLTM. Performance on these trials exceeded chance across all three presentations of each shape (all experiments, pAIC$\ll$0.001). These reports were not only statistically reliable but also highly accurate in absolute terms, with anchoring probability exceed-

ing 96% and precision exceeding 0.071(− 1) in all experiments (Table 1).

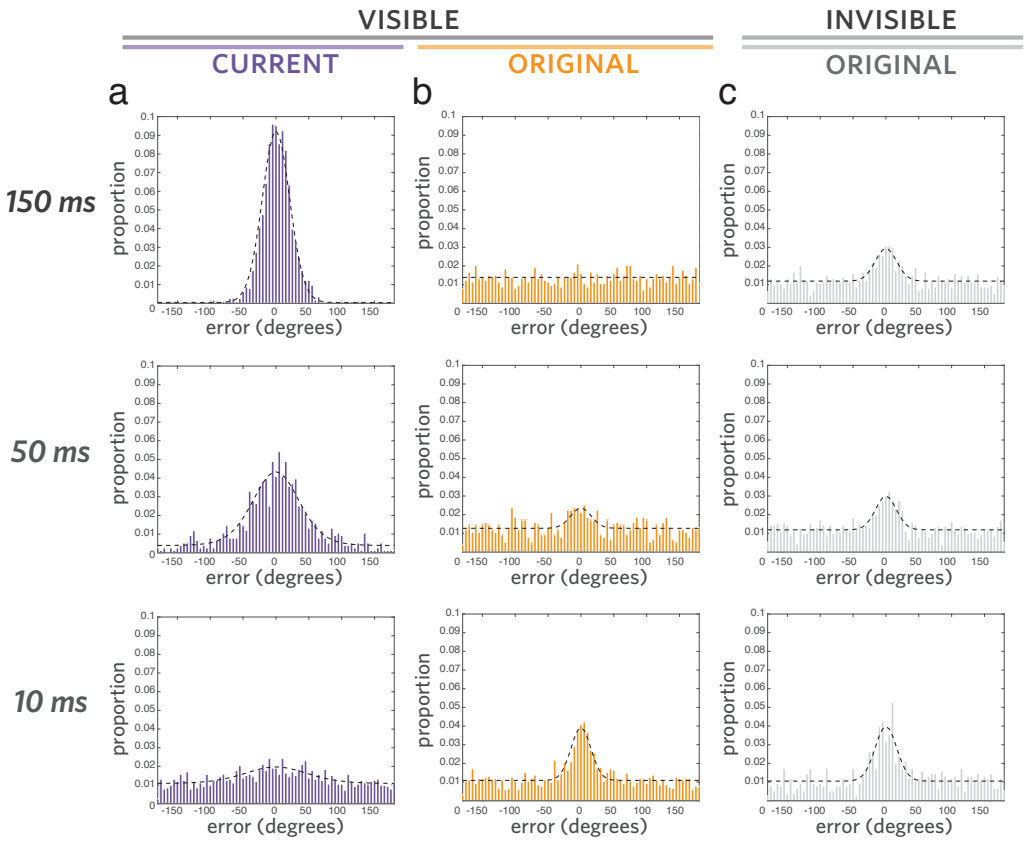


Figure 3.3: Distribution of Errors in Test Phase. Each panel contains a histogram of errors pooled across all observers and the corresponding fit from the three-component model (dotted line). In (a) and (b), all visible trials are plotted: in (a), errors represent the difference between each response and the current color, whereas in (b) errors represent the difference between each response and the original color. (c) On invisible trials, responses clustered around the original color.

3.2.2 Visible Test Trials

The task remained exactly the same on final test trials (Fig. 3), except that the shape (when visible) was presented for different durations (across experiments), was visually masked, and was rendered in a new, independently sampled color (θcurr). Performance on these trials

Table 3.2: Present/absent performance during the final test phase. Response criterion was calculated as in Snodgrass and Corwin (1988).

Experiment	Sensitivity	Criterion
E2 (50ms)	2.27	0.341
E3 (10ms)	0.753	-0.246

greatly exceeded chance (all experiments, pAIC<0.001). Confirming that the duration manipulation influenced performance, present/absent judgments in Experiment 3 (10 ms) were less sensitive (Table 2; t=6.3, p<0.001) and more conservative (t=4.4, p<0.001) than in Experiment 2 (50 ms).

We hypothesized that reducing the amount of sensory evidence for the current color would result in a poorer representation of the current color of the shape. Indeed, as stimulus durations decreased (150ms to 50ms to 10ms), so did anchoring proportion (slope: p<0.001; E1 vs. E2: p=0.002; E2 vs. E3: p=0.025) and precision (slope: p<0.001; E1 vs. E2: p<0.001; E2 vs. E3: p=0.003). This is consistent with the hypothesis that a decrease in the amount of sensory evidence leads to a concomitant increase in substitutions from VLTM (Fig. 4). Indeed, the probability that an observer's report was anchored on the original color increased as stimulus duration decreased (Table 3; slope: p=0.003; E1 vs. E2: p=0.006; E2 vs. E3: p=0.075). Duration did not, however, affect the precision of these reports from long-term memory (slope: p=0.215).

Table 3.3: Model parameter estimates for the current and original feature values when the stimulus was visible during final test phase.

Exp.	Anchoring Probability				Precision			
	Current		Original		Current		Original	
	MLE	*95% CI*	*MLE*	*95% CI*	*MLE*	*95% CI*	*MLE*	*95% CI*
E1 (150ms)	0.972	[0.900, 0.990]	0.000	[0.000, 0.032]	0.047	[0.043, 0.051]	0.028	[0.006, 0.080]
E2 (50ms)	0.729	[0.616, 0.848]	0.086	[0.021, 0.176]	0.025	[0.020, 0.031]	0.063	[0.022,0.084]
E3 (10ms)	0.251	[0.146, 0.685]	0.216	[0.124, 0.336]	0.017	[0.010, 0.024]	0.065	[0.038, 0.087]

3.2.3 Invisible Test Trials

No shape was presented at all (invisible or '0-ms' condition) on a subset of final test trials in all experiments, although visual masking disguised these trials from observers. We included these trials to isolate the contribution of VLTM to perceptual judgments, as no sensory information was available. Overall, performance on these trials exceeded chance (all experiments, pAIC<0.007). Intrusions occurred both on trials where observers erroneously reported that the shape had been present (pAIC<0.018) — thus misattributing their report to sensory input on the current trial — as well as on trials where they correctly reported that the shape had been absent (pAIC<0.031 Table 4). On these "present" trials, anchoring on the original color in VLTM was as likely (Experiment 3; p=0.276) or more likely (Experiment 2; p=0.030) than on "absent" trials. Neither anchoring probability nor precision differed across experiments (slope: ps>0.167).

3.2.4 Testing an alternative 'blend' model

The modeling approach so far assumes that a perceptual judgment entails sampling directly from either memory or sensory representations ('swap' mechanism), resulting in a bimodal mixture of response distributions that reflects the independent origins of the underlying signals. Another possibility, however, is that perceptual judgment entails sampling from an integrated representation ('blend' mechanism) that is formed by combining memory and sensory representations according to Bayes' rule, resulting a single unimodal response distribution (Knill & Pouget, 2004). Here we evaluate how well such a mechanism might explain our findings.

Formally, this Bayesian blend account asserts that responses are generated by sampling

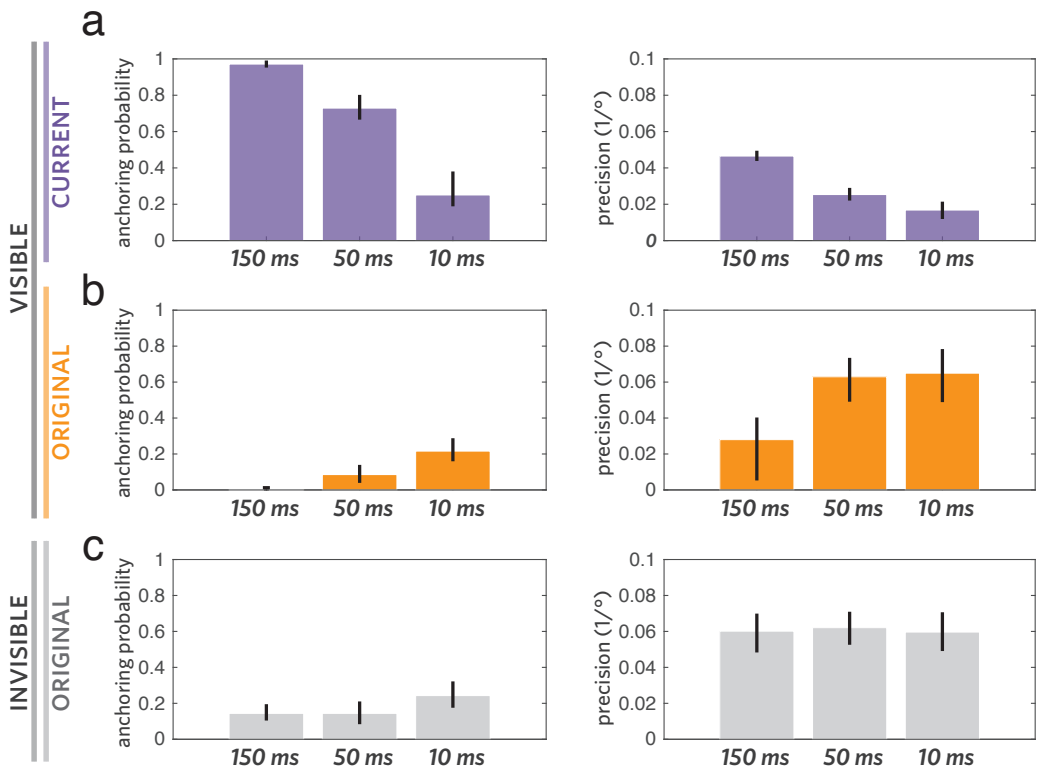


Figure 3.4: Modeling Results for Test Phase. (a) The probability of anchoring on the current color and the precision of these responses decreased as stimulus duration was reduced from 150ms to 50ms to 10ms. (b) This reduction increased the probability of anchoring on the original color. At shorter durations (50ms, 10ms), the precision of responses anchored on the original color was high, at least relative to the precision for the original color on initial exposure trials (see Table 1). (c) On trials containing no colored shape, anchoring on the original color occurred reliably, and at a consistent rate across experiments, with similar and high precision. Error bars represent bootstrapped 68% confidence intervals (comparable in width to 1 s.e.m).

Table 3.4: Model parameter estimates for responses anchored on the original feature value when the stimulus was invisible during final test phase (black text). Rows labeled 'present' with gray text contain parameter estimates based only on trials where the observer erroneously reported that the shape was present; 'absent' rows with gray text were based on trials where the shape was correctly reported as absent.

Experiment	Anchoring Probability		Precision	
	MLE	95% CI	MLE	95% CI
E1 (oms)	0.144	[0.080, 0.227]	0.060	[0.036,0.078]
E2 (oms)	0.144	[0.046, 0.263]	0.062	[0.044, 0.086]
"present" only	0.240	[0.093, 0.382]	0.056	[0.036,0.115]
"absent" only	0.119	[0.031, 0.242]	0.065	[0.0415,0.095]
E3 (oms)	0.243	[0.138, 0.384]	0.060	[0.042,0.084]
"present" only	0.323	[0.163, 0.499]	0.063	[0.046,0.102]
"absent" only	0.216	[0.102, 0.375]	0.057	[0.037,0.088]

from the posterior density function, which is proportional to the product of three functions: a likelihood function based on the original color seen during exposure trials, a second likelihood function based on current sensory information, and the prior density function representing the probability of the shape taking on any given color. This latter probability is uniform across colors, so it does not affect the proportionality relationship:

$$p(\hat{\theta}|\theta_{orig}, \theta_{curr}) \propto p(\theta_{orig}|\hat{\theta})p(\theta_{curr}|\hat{\theta}) \tag{3.3}$$

We again assume that the current and original color representations follow von Mises distributions, each centered on the correct feature value θ_{curr} and θ_{orig}, respectively, with concentration parameters κcurr and κorig, capturing the amount of uncertainty in the representation of the current color and noise in the long-term memory representation, respectively.

The product of these two von Mises distributions is also a von Mises distribution centered on θblend with concentration parameter κblend, given by:

$$\theta_{blend} = atan2(\kappa_{orig}\sin(\theta_{orig}) + \kappa_{curr}\sin\theta_{curr}, \kappa_{orig}\cos(\theta_{orig}) + \kappa_{curr}\cos\theta_{curr}) \quad (3.4)$$

$$\kappa_{blend} = \sqrt{\kappa_{orig}^2 + \kappa_{curr}^2 + 2\kappa_{orig}\kappa_{curr}\cos(\theta_{curr} - \theta_{orig})} \quad (3.5)$$

We used maximum-likelihood estimation to derive estimates of κcurr and κorig that provided the best fit to errors on visible test trials across all three experiments, then converted these κ values to σ-1 values in units of degrees for ease of interpretation. To better account for large errors, a weight parameter β was also estimated, reflecting the contribution of this integrated representation relative to a uniform 'guessing' distribution:

$$p(\hat{\theta}) = \beta\phi_{\sigma_{blend}}(\hat{\theta} - \theta_{blend}) + (1-\beta)\frac{1}{360} \quad (3.6)$$

Interestingly, the resulting parameter estimates for this blend model suggest an overall interpretation that is compatible with those based on the swap model, namely that the contribution of the current color representation to the posterior diminishes (i.e., its precision decreases) as presentation duration is reduced, while the contribution of the original color representation increases (Table 5). These trends are accompanied by increasing guessing proportions, possibly reflecting greater overall response uncertainty at shorter durations. Nevertheless, we found that the blend model consistently underperformed relative to the swap model, as quantified by the difference in AIC (Table 5 and Supplemental Figure 1). Thus,

while the blend model yields interpretable parameter estimates, the data suggest that the swap mechanism is more likely.

Consistent with these results, visual inspection of normalized error distributions (Figure 5) in the 50-ms and 10-ms conditions suggests that they are bimodal, with modes centered on the current and original colors. This is inconsistent with the blend account, which, critically, predicts a unimodal distribution that peaks between these two color values. Taken together, these findings lend further support to the idea that VLTM manifests in perceptual judgments as discrete substitutions, as expressed by the swap model.

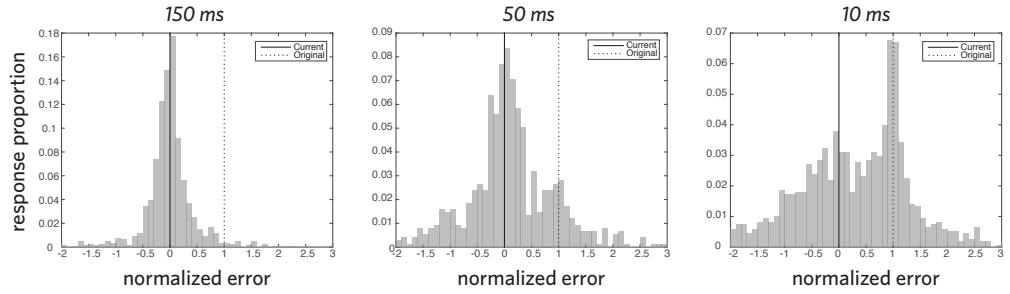


Figure 3.5: Distribution of responses relative to the current and original colors on visible test trials. Errors were normalized by taking the difference between the response and current color and dividing it by the distance between the original color and current color for each item. This results in a distribution in which the current and original colors have been remapped to 0 and 1, respectively, and errors have been scaled in proportion.

One caveat we would like to add to these model-comparison results is that the AIC differences per participant are small (range: 0.24-3.03), as each model was fit to data aggregated across 24 participants. As a rule of thumb, AIC differences smaller than 5 should be interpreted with caution (Burnham & Anderson, 2004). Thus, we consider the blend account to still be very theoretically interesting and potentially relevant. We speculate that blend and swap mechanisms need not be mutually exclusive, and that some version of each may dominate in different situations. Specifically, large and random disparities between old and new

Table 3.5: Blend-model parameter estimates for visible trials during the test phase. For side-by-side model comparison, the final column reports the corresponding AIC for the swap model.

Exp.	Anchoring Prob.	Orig. Precision	Curr. Precision	AIC_{blend}	AIC_{swap}
E1 (150ms)	0.965	0.011	0.047	10451.80	10446.05
E2 (50ms)	0.727	0.011	0.025	8309.03	8279.02
E3 (10ms)	0.267	0.067	0.014	17584.33	17511.62

information (as in the current study) may promote the formation of more discrete representations, leading to memory substitution errors, whereas an alternative scenario with smaller and/or predictable disparities between old and new information may encourage integration of these two sources of information, leading primarily to memory-driven biases at test (Golomb et al., 2014).

In addition, there exists an interesting intermediate possibility between these strict blend and swap mechanisms. The current formulation of the Bayesian 'blend' account assumes that the two likelihood functions reflecting VLTM and sensory information are von Mises distributed, but guessing nevertheless occurs on some proportion of trials. If it were the case that failure to retrieve the color association in VLTM or failure to register current sensory information sometimes occurred with independent probabilities, then the posterior distribution would be formed by combining mixtures of von Mises and uniform distributions by multiplication, according to Bayes' rule. When the disparity between original and current colors is large enough, the resulting posterior distribution is bimodal and very similar in form to the mixture distribution predicted by the swap model, except that the modes are slightly shifted towards each other. This alternative model is intriguing because it provides a subtly distinct account of the source of the bimodality in response distributions — as arising

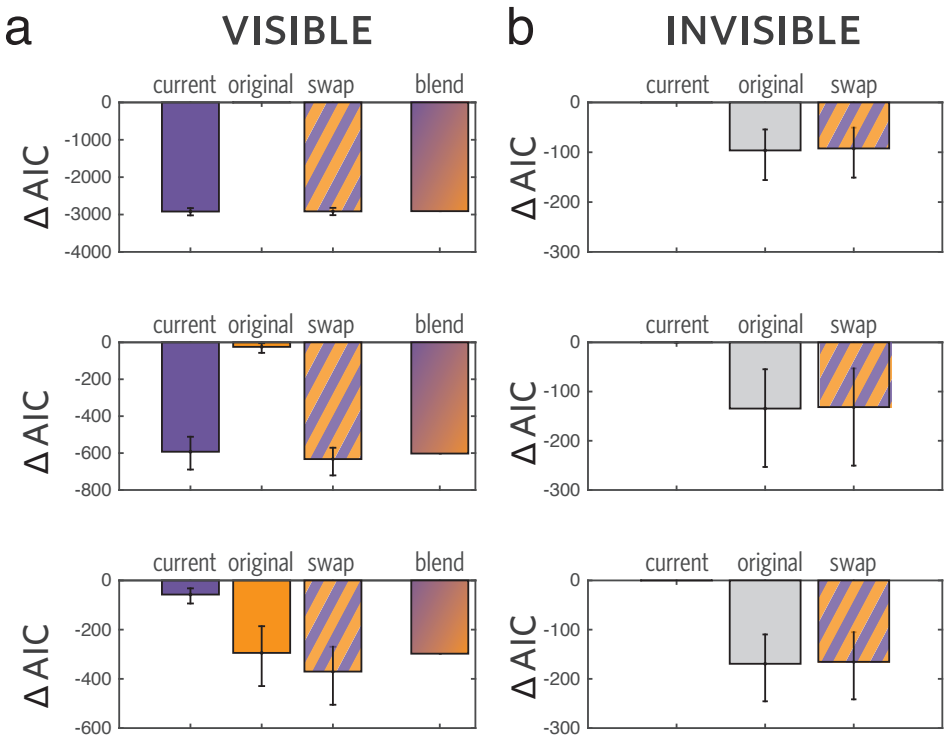


Figure 3.6: Model Comparison Results. (a) On visible test trials, the three-component mixture model ("swap") provided the best overall fit to the data across experiments, relative to reduced models containing only the original-color component or only the current-color component, plus the uniform component. Differences in AIC are plotted relative to the null model containing only the uniform component. The blend model, while performing well, did not exceed the swap model in any experiment. b) On invisible test trials, where there was no current color, the two-component model containing the original-color and uniform components significantly outperformed the null model. The performance of the swap model was comparable, except for the penalty due to the additional parameter. Error bars indicate 68% confidence intervals.

from direct updating of a single underlying representation that supports two feature values vs. arising from sampling a mixture of two representations, each supporting one feature value. However, it is several times more computationally intensive to carry out the necessary simulations to compare these two accounts, since the product of VM+U mixtures does not yield parameters with closed-form definitions. Moreover, the large, random disparities used in the current study may make such small biases especially difficult to detect. Nevertheless, although the current study design was not optimized to resolve differences between these two accounts, we hope that future investigations will employ designs that are more sensitive to measuring small biases in bimodal distributions.

3.3 Discussion

The present study aimed to elucidate how object representations in long-term memory and sensory information about a just-seen object mutually inform perceptual judgments. We examined this question by associating colors with shapes in VLTM, and then testing how these memories informed judgments about the color of these shapes when they later appeared again. We found independent contributions of both long-term memory and immediate sensory input on responses, with a greater likelihood of drawing upon VLTM as sensory information decreased. This study provided a strong test of the contribution of VLTM, as these memories were both task-irrelevant (observers were instructed to judge the current color) and orthogonal to the current information (original and current colors were uncorrelated). Interestingly, the precision of these memory substitutions was comparable to the precision with which colors were initially encoded (see also Brady, Konkle, Gill, Oliva, & Alvarez, 2013), and was not modulated by the amount of sensory input. Although we interpret

these findings as reflecting automatic substitutions from VLTM during perceptual judgment, an alternative possibility is that observers employed a controlled strategy of defaulting to the original color on certain trials, such as when they did not see the shape. However, this account is insufficient in several ways: (1) the original color was objectively uncorrelated with the current color across shapes, making it irrelevant to the task of reporting the current color; (2) observers accurately reported the current color when the shape did appear on test trials, showing that they did not assume that the original color was the correct response on these trials; (3) anchoring probabilities on the original color were reliable but notably smaller (8%-22%) than in explicit long-term memory tests in similar paradigms (40%-65%; (Brady et al., 2013, Fan & Turk-Browne, 2013a), suggesting that anchoring did not simply reflect the accessibility of representations in VLTM; (4) observers reliably responded that a shape had been "present" after reporting the original color on invisible trials, thus misattributing their report to sensory input on the current trial; and (5) on these "present" trials, anchoring on the original color was as likely (Experiment 3; p=0.276) or more likely (Experiment 2; p=0.030) than on "absent" trials. Together, these points are more consistent with automatic than controlled expression of VLTM during perceptual judgment. Specifically, the representation of the original color is reactivated when there is sensory overlap with specific prior experiences (i.e., a matching shape cue), and this reactivation from memory may be falsely attributed to immediate sensory experience, thereby influencing perceptual judgments even when this influence confers no benefit to (and perhaps even impairs) task performance. In this study, we collected present/absent judgments in order to assess the relationship between subjective judgments about the visibility of the current shape and color responses. A promising direction for future research would be to also measure the relation-

ship between subjective confidence in the precision of the current color representation and the rate of substitutions from visual long-term memory, which would provide insight into the metacognitive processes that arbitrate between sensory and long-term representations on a moment-to-moment basis (Suchow et al., 2012).

Throughout, we employed a mixture-modeling approach to gain basic insights into interactions between long-term memory and sensory information in perceptual judgments. In future studies, alternative modeling techniques, in which encoding precision is modeled as varying across items and trials (van den Berg et al., 2012; Fougnie, Suchow, & Alvarez, 2012) and/or determined by variability in spiking activity of neural populations (Bays, 2014), may yield a more nuanced understanding of how such interactions influence the content and fidelity of visual representations. Taken as a whole, our findings suggest that the visual system automatically recruits long-term memory during perceptual decision-making. Ultimately, such retrieval mechanisms may be crucial for deploying past experience efficiently and for improving perceptual inference.

4

What kinds of feedback enhance perceptual learning?

It is commonly observed that perception and memory can be used to guide our actions and decisions. However, our actions also affect what we see and remember (Gureckis

& Markant, 2012), such as when we actively explore a new city, making choices about where to go and what to look at. Similarly, practicing certain actions can lead to dramatic improvement in our skill in performing them, such as when we invest time in practicing a musical instrument. How do humans intelligently select actions to best achieve their goals? My research on this topic examines the contribution of task practice and performance feedback to learning of perceptual skills, taking the phenomenon of 'statistical summary perception' as a case study.

4.0.1 How does sensory feedback guide learning?

We often encounter objects in groups: fallen leaves on the lawn, food and beverages on a dining tray, or a stack of papers on our desk. In such cases, we may have the goal of interacting effectively with the group of objects: minimizing time raking up the leaves, keeping the tray balanced while carrying it across the room, or moving the stack of papers without tipping it over. Achieving these goals relies upon accurately representing not only the individual objects that comprise each group, but general properties of the whole group of objects, such as the area of highest density or its center-of-mass.

Prior investigations have shown that people are able to quickly and accurately extract such "statistical summary features" from groups of objects. For example, naive participants can extract the mean size of an array of disks of varying sizes (Chong & Treisman, 2003a, 2005), the mean orientation of peripherally-presented oriented gratings (Parkes et al., 2001), and the mean emotional expression of a set of faces (Haberman & Whitney, 2009). Good performance on these summary judgment tasks has been found even when participants are highly impaired at judging the properties of an individual object in the set (Ariely, 2001).

However, although these findings show that certain summary features are readily accessible, the computations underlying summary perception are not fully understood (Bauer, 2009, Myczek & Simons, 2008, Ariely, 2008). Moreover, these demonstrations do not explain how people acquire this ability, nor explore conditions for further learning.

Examining how learning affects statistical summary perception may provide crucial insight into how these summary representations are formed and used to guide behavior. Generally speaking, such learning may proceed in two ways: (1) in an unsupervised manner, in which mere exposure to sensory inputs leads to better discriminability along task-relevant dimensions; or (2) in a supervised manner, in which external feedback about the correctness of discrimination judgments during training improves accuracy in other contexts. A major goal of the current study was to directly examine the contribution of feedback to learning in statistical summary perception (supervised), especially as compared to the effects of experience alone (unsupervised).

Perceptual learning entails the enhancement of performance on a perceptual task (e.g., orientation, size, position discrimination) due to practice. The role of external feedback in perceptual learning, however, is complex. While several studies report little effect of feedback on performance improvements (Ball & Sekuler, 1987, Karni & Sagi, 1991a, Shiu & Pashler, 1992), others have found that feedback enhances and may even be necessary for perceptual learning (Herzog & Fahle, 1997b, Seitz et al., 2006). One possible reason for this lack of consensus is that feedback has largely been treated in a unitary fashion. Specifically, traditional experimental procedures used to examine perceptual learning have relied on binary judgments, and have thus been restricted to manipulating feedback in a purely success-based fashion (Liu et al., 2014).

In everyday life, however, where people not only look at sets of objects, but also interact with them, sources of feedback other than binary task success are available. Specifically, while interacting with objects, we continuously experience sensory feedback arising from vision and proprioception. For example, upon lifting an unevenly loaded tray, we quickly notice if the tray is beginning to tilt, allowing us to compensate for unexpected forces by adjusting the position of our grip in a continuous manner. The mismatch between the expected position of the tray upon lifting it (balanced) and its actual position (tilted) constitutes a form of sensory-prediction error, which conveys visual and proprioceptive information about the deviation between the expected and actual consequences of an action (Synofzik et al., 2006). Sensory-prediction errors have been shown to be crucial for basic kinds of motor learning (Shadmehr et al., 2010). For example, when reaching towards a target location, the mismatch between the expected hand position based on the motor command and actual hand position based on visual information is sufficient to induce adaptation of subsequent reaching movements to account for earlier errors (Tseng et al., 2007). More generally, the amount and kind of sensory feedback available during training can directly affect the nature of learning and the extent of generalization of motor performance to new contexts (Izawa & Shadmehr, 2011a, Nikooyan & Ahmed, 2015, Taylor et al., 2012). Finally, an active movement appears to be necessary for sensorimotor adaptation, as passive viewing of errors does not produce sensory-prediction errors (Held & Gottlieb, 1958, Kording & Wolpert, 2004a).

In order to directly examine the contributions of these different forms of error feedback to statistical summary perception, we developed a reaching task requiring participants to make pointing movements towards the centroid (i.e., average location) of an ar-

ray of dots (Fig. 4.1). We hypothesized that providing vector error feedback (i.e., conveying both distance and direction error) after the pointing movement would generate sensory-prediction errors, and lead to improved perceptual discrimination of the centroid after training. Whether other forms of feedback conveying less spatially precise error information (i.e., reward feedback conveying only distance to the centroid) might improve perceptual discrimination of the centroid was an open question. Before and after training, we tested changes in discrimination ability using an independent perceptual test.

4.1 Experiment 1: Centroid Discrimination

The goal of this experiment was to investigate the influence of error-related feedback on the fidelity of the perceptual representation of the centroid of a set of visual objects. We tested this by manipulating whether participants received vector feedback or scalar feedback on a trial-by-trial basis while practicing a task requiring pointing movements towards the centroid. To provide a baseline measurement of the effects of task practice alone, we also included a control condition in which participants did not receive any trial-by-trial feedback.

We hypothesized that vector feedback might lead to faster improvement during training than the other conditions, so we planned to examine the time course of performance on the pointing task. In order to distinguish between task-related practice effects from genuine changes in the fidelity of the underlying representation, we also measured the degree of improvement on a separate perceptual task performed before and after pointing-task training. This perceptual task was designed to match the pointing task insofar as it required computing the centroid location, although it required a different response.

4.1.1 Methods

Participants

Forty-five young adults (29 women, age range: 18-34y) were recruited for participation from the research subject pool at Princeton University or the Princeton, NJ community. In this and all subsequent experiments, each participant received course credit or $24 for participation and all provided informed consent to a protocol approved by the Princeton IRB. All participants were right-handed and reported normal or corrected-to-normal visual acuity.

Apparatus

The experiment was controlled by a computer running MATLAB 2013a and Psychtoolbox (http://psychtoolbox.org). Stimuli were presented on a vertically oriented 58.42-cm touchscreen display (Acer T232HL LCD) with 60 Hz refresh rate positioned 43 cm from the participant.

Stimuli and procedure

Stimuli were presented against a dark gray background on a touchscreen display (49.9° x 40.0°; Fig. 2A). In all phases, trials began with the onset of a central fixation cross, presented for 550–1000ms (randomly jittered). The stimulus array comprised 8 white dots (diameter=0.9°) presented for 200 ms. This number of items was chosen since it exceeds traditional estimates of the capacity of visual short-term memory (Luck & Vogel, 1997, Zhang & Luck, 2008), discouraging explicit encoding and averaging strategies, while still being small enough that our measures would remain sensitive to errors in perceptual averaging. The stimulus

duration was chosen in order to match those used in earlier studies of statistical summary perception (Chong & Treisman, 2003b) and be short enough to prevent multiple eye movements within a trial (Saslow, 1967). Following response, the display remained blank for 1500ms. Individual dot locations were independently sampled from a bivariate Gaussian distribution (circularly symmetric; σ of marginal distributions=11.2°). Specifically, on each trial we first sampled 8 locations from this bivariate Gaussian distribution centered on the origin. Because the mean of a random sample is likely to deviate from the true mean of the underlying distribution, we next applied a rigid spatial shift to all 8 locations so that the sample centroid would exactly coincide with the origin. Finally, in order to generate the final display, we applied another rigid spatial shift to all 8 locations in order to move the sample centroid to the predetermined centroid location for that trial.

The centroid is defined as the average of the position coordinates of all dots. More precisely, the coordinates of the centroid, xc and yc, is computed as the sum of the product of the individual areas and coordinates of each dot, divided by the total number of all dots on the display:

$$x_c = \frac{1}{n} \sum x_i \tag{4.1}$$

$$y_c = \frac{1}{n} \sum y_i \tag{4.2}$$

The experimental protocol consisted of three phases over two consecutive days: a pre-test phase on Day 1, and training and post-test phases on Day 2. The post-test was administered immediately after the conclusion of the training phase.

Quadrant task. The pre-test and post-test phases consisted of a four alternative forced choice (4AFC) perceptual task (4 blocks, 50 trials/block), which entailed discriminating which of the four display quadrants (NE, NW, SE, SW) contained the centroid and tapping the display anywhere within that quadrant of the display after stimulus offset with their right finger (Fig. 2A). Response time (RT) was measured as the time elapsed between stimulus offset and response (Table 4.1).

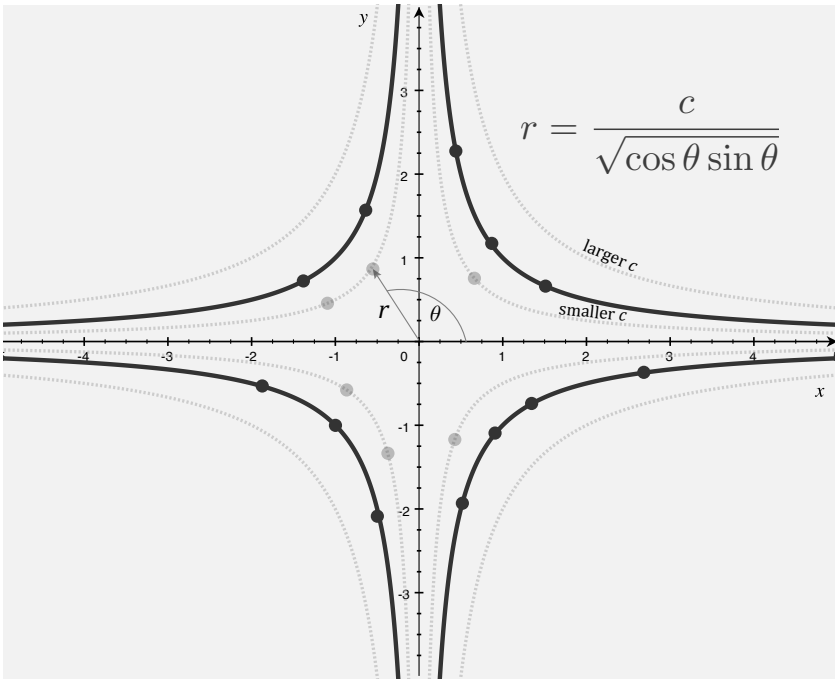


Figure 4.1: Figure 1: Distribution of centroid locations in the quadrant task followed a hyperbolic function: $r = \frac{c}{\sqrt{\cos(\theta)\,\sin(\theta)}}$, where r is the radial distance from the center of the display, □ represents randomly sampled angles, and c is a scaling factor that controls the overall distance to the nearest quadrant boundaries. Smaller values of c produce displays with centroids falling closer to boundaries, thus making discrimination more difficult.

Given the long history of using discrimination judgments to assay the quality of perceptual representations in visual psychophysics (Herzog & Fahle, 1997a, Petrov et al., 2005), we thought such a task might be appropriate. In traditional perceptual learning tasks that entail

discriminating perceptual features that differ along a single dimension (e.g., orientation), binary judgments are appropriate. Because the location of the centroid varied along two dimensions (i.e., horizontal and vertical position), we reasoned that the minimal space of response options would be four alternatives. Thus, dividing the display into four quadrants was the most natural way of achieving this.

In order to make the task more consistently difficult, we constrained the location of dots to appear close to the quadrant boundaries (Fig. 1). Specifically, centroid locations in the quadrant task were distributed following a hyperbolic function:

$$r = \frac{c}{\sqrt{\cos(\theta)\sin(\theta)}} \tag{4.3}$$

, where r is the radial distance from the center of the display, θ represents randomly sampled angles, and c is a scaling factor that controls the overall distance to the nearest quadrant boundaries. Smaller values of c produce displays with centroids falling closer to boundaries, thus making discrimination more difficult. The rationale for using the hyperbolic function derives from the spatial layout of the four quadrants on the display. Specifically, we assume that the likelihood that the participant will select the correct quadrant varies according to the angular location of the centroid. At angular locations near the main diagonals (y = x and y = -x), the distance between the centroid and each quadrant boundary is more closely matched. However, as the centroid approaches angular locations near the vertical (y-axis) or horizontal (x-axis) meridians, its relative distance to the boundary dividing two adjacent quadrants is much smaller than its relative distance to the other two quadrants, effectively biasing discrimination towards a binary choice between the correct and nearest quadrant, rather than a 4-way choice among all quadrants. A hyperbolic function was chosen in order

to better match discrimination difficulty at all angular locations (at least relative to a circular function, where radial distance is constant at all angles). Nevertheless, because centroid locations were sampled from the same distribution in all conditions, the specific choice of distribution does not bear on the most critical comparisons in the current study, which are between conditions. In order to ensure that the difficulty of trials in the pre-test and post-test phases was equated, we used a separate thresholding procedure on Day 1 to estimate the value of c that would result in baseline accuracy of 62.5% on the 4AFC task (QUEST; (Watson & Pelli, 1983). No feedback was given during the thresholding, pre-test and post-test phases. We restricted values of c to fall in a broad range: [0.344 cm, 3.44 cm]. In the rare case that estimated values of c fell out of this range, we extended the thresholding phase until c converged on an acceptable value. Nevertheless, estimation of c only served to calibrate the difficulty of the quadrant task such that participants achieved approximately 62.5% accuracy on the task (the midpoint between 25% and 100%). Because the value of c was fixed within-participant, changes in perceptual discrimination were measured by computing the percent change in accuracy between the pre-test and post-test phases for each participant.

Pointing task. The training phase consisted of a centroid estimation task (12 blocks x 50 trials/block = 600 total trials), which entailed tapping the display at the estimated location of the centroid with the right index finger (Fig. 2A). Response time was measured the same way as in the quadrant task (Table 4.1).

Centroid locations in this pointing task were sampled from a uniform distribution over a rectangular region spanning the central 25% of the display. Note that the distribution of centroid locations in the pointing task differs from that used in the quadrant task in this and all experiments. The type of feedback provided to participants was manipulated across sepa-

rate groups during the training phase. Participants in the Vector condition (N = 15) received trial-by-trial error information about the distance and direction of the actual centroid from their response. This vector error was provided by marking the participant's response with a blue crosshair and the actual centroid with a green crosshair. Participants in the Scalar condition (N = 15) received trial-by-trial error information proportional to the distance of their response from the centroid but no direction-error information. Scalar error was returned as a numerical point score (range=0-10 points), where the number of points earned followed a Gaussian reward function of deviation from the centroid (with $\sigma = \frac{11.2}{\sqrt{8}}$), rounded to the nearest integer value. Participants in the Control condition (N = 15) did not receive error feedback on a trial-by-trial basis. Regardless of condition, all participants received task performance feedback at the end of each 50-trial block in the form of a running cumulative point total and participants were encouraged to try to earn as many points as possible. To familiarize participants with the pointing task, participants completed 8 practice trials without feedback.

Data analysis. On each trial, error was computed as the distance between the actual centroid and participant's response. Outlier responses with errors exceeding 3x the standard deviation of each participant's errors were removed prior to computing aggregate statistics of performance. Overall performance for each participant was quantified as root-mean-squared error (RMSE). RMSE is a summary statistic that captures overall deviation between the centroid and participants' responses, and reflects both the accuracy (bias) and precision (variability) of responses. We also report the mean radial error (average distance between the centroid and response) and the standard deviation of errors, computed within-participant, to assess potential dissociations between these two error components.

While the primary goal of this study was to examine the consequences of training with feedback on subsequent perceptual discrimination of the centroid, we also examined changes in performance over the course of the training phase, reasoning that such changes might help explain any training effects observed. The time course of errors was fit with an exponential function ($error_{abs} = be^{ax}$) for each participant, where the exponential parameter a captures the rate of change in performance over the course of training (where positive values indicate increasing error and negative values indicate decreasing error). The same exponential function was also fit to response times during the training phase.

We also examined the presence of any directional biases (e.g., consistently pointing above and to the right of the centroid) by measuring the angular component of the error vector. These angles were tested for uniformity using the Hodges-Ajne test, a non-parametric test of deviations from uniformity in circular data (Ajne, 1968; Zar, 1999). In the absence of a directional bias, the distribution of these angles should be circularly uniform around the centroid on average (in the range [-180°,180°]), where θ=0 was defined as pure rightward bias, and counterclockwise rotation is positive). In order to derive an empirical estimate of chance performance on this task, we conducted simulations informed by the actual spatial distribution of responses participants made in this task (as opposed to assuming uniform probability of responding at all locations in the display). Specifically, for each trial in the pointing task, we simulated random responding by sampling from the set of responses a given participant had made, and computed error on the basis of this random response. This procedure was repeated 1000 times for each participant in order to derive robust estimates of RMSE under completely random responding. Alternatively, participants, although not guessing completely at random, may have been able to achieve above chance-level perfor-

mance by subsampling the display (e.g., pointing to the average location of only 3 or 4 dots, rather than the full set of 8). This possibility was examined by estimating the goodness-of-fit (RMSE) of each of the intermediate cases between random guessing (i.e., 0 of 8 dots) and full averaging of all dots presented (i.e., 8 of 8 dots) to each participant's actual response distribution. Specifically, this was estimated by sampling a random subset of dots from each trial of a fixed size (e.g., 3 of 8 dots), recording the distance between the centroid of this random subset and the participant's response, and then computing RMSE. This procedure was performed 1000 times for each subset size, for each participant, in order to derive robust estimates of average deviation between participant responses and the centroids of these random subsets. Although dot locations were sampled from the same underlying distribution, individual trials varied in how dispersed the dots were on the display. Trials with high dispersion might be more difficult (i.e., result in less precise centroid representation) than trials with low dispersion. To quantify how dot dispersion affected centroid estimation, we used multiple regression to model the contribution of horizontal location variability (sx) and vertical location variability (sy) to error magnitude across trials in the pointing task for each participant ($error = s_{xi}^T \beta_{sx} + s_{yi}^T \beta_{sy} + \varepsilon_i$).

4.1.2 Results

Pre-test

Pre-test performance on the quadrant task in all groups was significantly better than chance (59.0% correct vs. 25%; $t(44)$=25.6 $p \ll 0.001$, d=3.81), and did not differ among groups ($F(2,42)$=1.76, p=0.184, $\eta 2$=0.078). Estimates of c, which controlled quadrant task difficulty (see Methods), also did not differ between groups at baseline (Vector: 0.958 cm [s.e.m. =

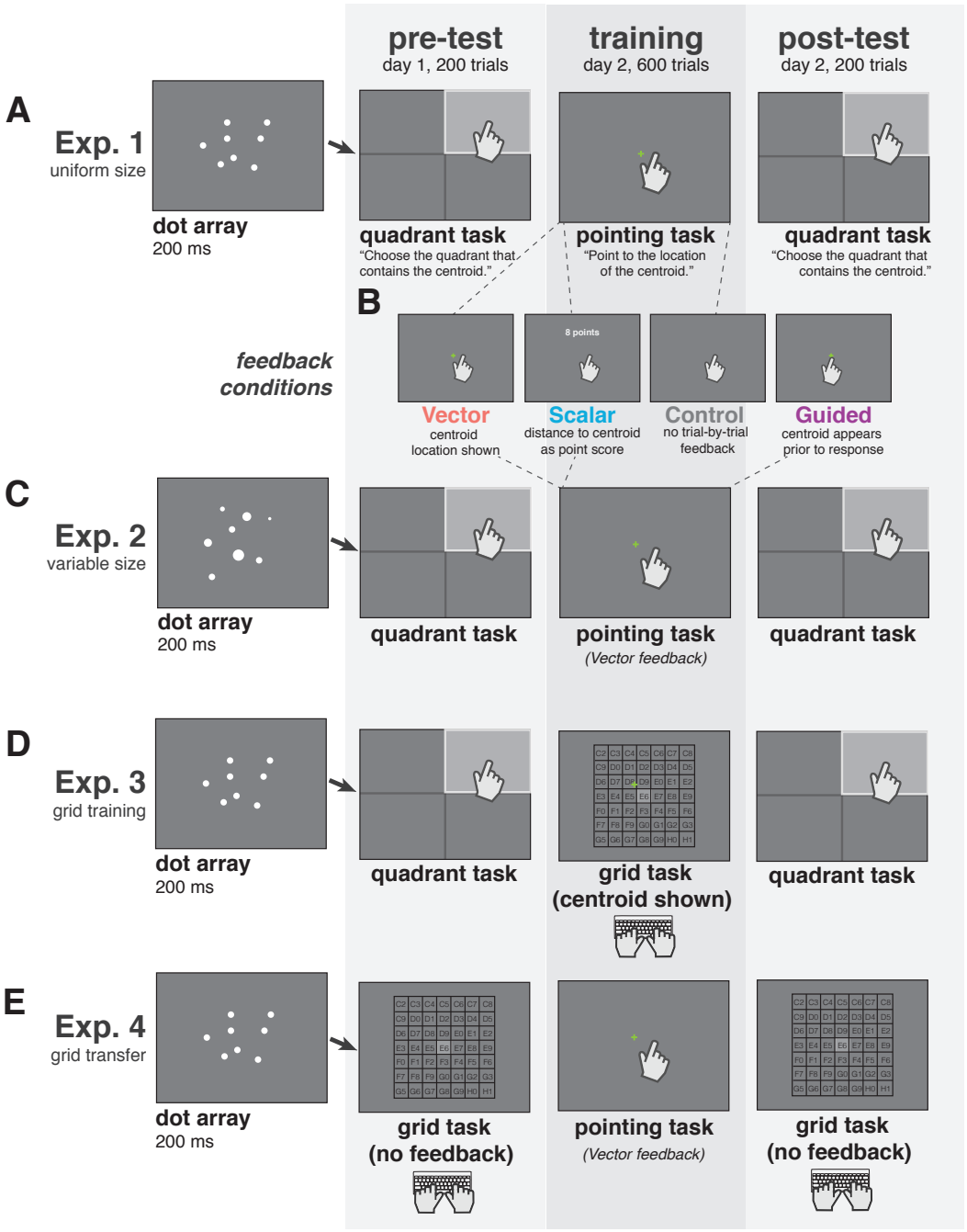


Figure 4.2: Task and procedure. (A) Eight dots of uniform size presented for 200ms in random locations on each trial. During training, participants pointed to the centroid on touchscreen. During the pre-test and post-test, participants discriminated which quadrant contained the centroid. (B) Groups received different trial-by-trial feedback during training: true centroid location shown after response (Vector), point score reflecting distance to centroid (Scalar), no feedback (Control), or centroid location shown prior to responding (Guided). (C) In Experiment 2, dots appeared in random sizes and locations. (D) During training in Experiment 3, participants typed in the code of the grid cell containing the centroid, and then were presented with the location of the centroid. (G) Participants in Experiment 4 performed the pointing task during training while receiving vector error feedback; the grid task was used to obtain a more precise measure of changes in perceptual discrimination.

0.123 cm], Scalar: 1.01 cm [s.e.m. = 0.097 cm], Control: 1.21 cm [s.e.m. = 0.194 cm]; $F(2,42)=0.814$, $p=0.450$, $\eta 2=0.0373$), nor did median response time ($F(2,42)=1.46$, $p=0.244$, $\eta 2=0.065$).

Training

Overall. Participants received different kinds of information following a response in the pointing task during the training phase (Fig. 2B). Performance in all groups highly exceeded that expected by chance ($p \ll 0.001$), verifying that participants were engaged with the task (Fig. 3A). Overall performance (RMSE) during the training phase did not reliably differ among groups ($F(2,42)=1.76$, $p=0.184$, $\eta 2=0.078$; Vector RMSE: 1.35cm; Scalar RMSE: 1.41 cm; Control RMSE: 1.54 cm). Radial error (i.e., reflecting the accuracy component of performance) also did not differ among conditions (Vector: 1.19 cm, Scalar: 1.24 cm, Control: 1.35 cm; $F(2,42)=1.61$, $p=0.212$, $\eta 2=0.071$; Fig. 3B); nor did error variability (i.e., reflecting the precision component of performance; Vector: 0.636 cm, Scalar: 0.678 cm, Control: 0.748 cm; $F(2,42)=2.17$, $p=0.127$, $\eta 2=0.094$).□

Time course. We thought it reasonable to expect participants to show steady improvement over the course of training, where the rate of improvement would be highest for the groups receiving either Vector or Scalar feedback. Surprisingly, these time series were generally flat (i.e., non-significant exponential learning rate) over the training period tested in all three groups (Vector: $t(14)= 0.996$, $p= 0.336$; $d= 0.257$; Scalar: $t(14)= 0.905$, $p= 0.381$, $d= 0.233$; Control: $t(14)= 0.676$, $p= 0.510$, $d=0.175$), suggesting that participants may have been performing close to ceiling. Although response accuracy did not improve over time, response times did significantly decrease over the course of training in all three conditions (Vector: $t(14)= 2.27$, $p=0.040$; $d= 0.585$; Scalar: $t(14)= 4.52$, $p<0.001$; $d=1.17$; Control: $t(14)=$

4.92, p<0.001; d=1.27). Neither the rate of decrease in RTs (F(2,42)=1.12, p=0.335, η2=0.051) nor overall RTs (F(2,42)=0.488, p=0.617, η2=0.023) differed between conditions, suggesting that this temporal trend in RT more likely reflects generic task practice effects, rather than the consequences of differences between feedback conditions.

Directional bias. Despite the lack of overall RMSE effects during training, there may have been differences in the degree to which participants were biased to respond between conditions. Using the Hodges-Ajne test for non-uniformity in the angular component of errors, we found significant evidence for directional bias Control participants (p=0.012; 95% CI: [−23.9°, −90.0°]): a bias below and to the right of the centroid. There was a trend toward deviation from uniformity in the Scalar group (p=0.071; 95% CI: [5.1°, −77.6°]), and no reliable non-uniformity in Vector participants (p=0.250; 95% CI: [−22.9°, 109.6°]). Nevertheless, 95% confidence intervals for the direction of bias overlapped heavily in the quadrant lying southeast of the centroid for all groups. These results are consistent with the possibility that there may be a southeast bias in the pointing task, which was reduced for participants in the Vector condition who received spatially precise directional error information on every trial.

Relationship to stimulus variability. While the locations of dots were drawn from the same distribution, their exact locations varied from trial-to-trial. We reasoned that trials with higher variability in dot location would be more difficult, and thus be associated with larger errors on average. To test this possibility, we used multiple regression to model the contribution of horizontal location variability and vertical location variability to error magnitude across trials for each participant. Across all conditions, we found that both horizontal location variability and vertical location variability were both highly predictive of error magnitude (Vector: β_{sx}=0.056: t(14)=8.48, p<0.001; β_{sy}=0.028: t(14)=3.19, p=0.007;

Scalar: β_{sx}=0.083: $t(14)$=6.57, p<001; β_{sy}=0.068: $t(14)$=6.66, p<0.001; Control: β_{sx}=0.084: $t(14)$=7.27, p<0.001; β_{sy}=0.076: $t(14)$=7.75, p<0.001). To examine potential differences between conditions due to feedback, we submitted these linear coefficients to a 2 axis (horizontal, vertical) x 3-condition (Vector, Scalar, Control) ANOVA. We found that the degree of dependence between variability and error was lower for Vector participants than for Scalar or Control participants (main effect of condition: $F(2,84)$=8.45, p<0.001, $\eta 2$ =0.167), suggesting that vector error feedback may serve to reduce noise in centroid estimation for more difficult trials.

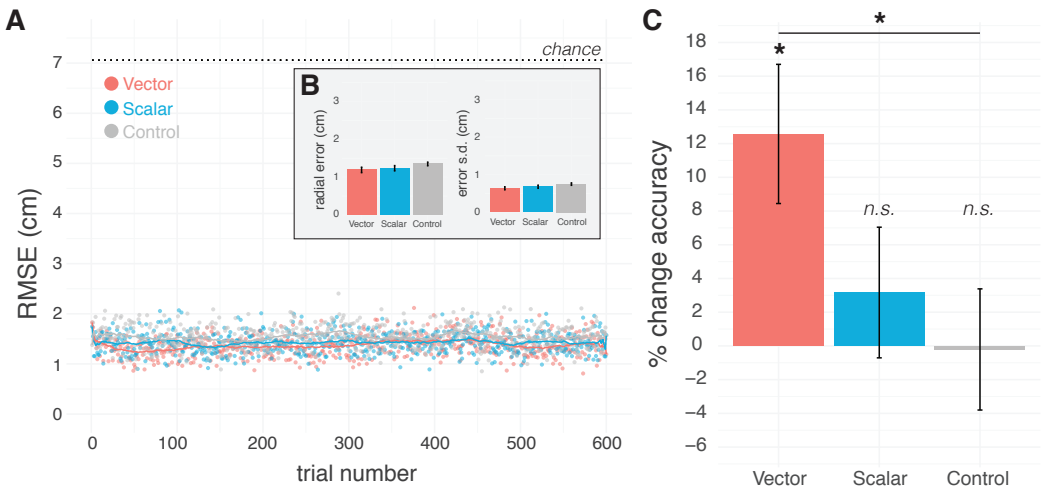


Figure 4.3: Experiment 1 Results. A) Overall performance (RMSE) on the pointing task during the training phase as a function of trial number for the Vector (Salmon), Scalar (Blue), and Control (Gray) groups. All groups performed better than chance (dashed line). B) Performance in the pointing task broken down into accuracy (radial error) and precision (standard deviation). C) Percent change in accuracy between post-test and pre-test in the quadrant task. Error bars represent 1 SEM *<0.05, <0.1.

Post-test

Despite showing similar performance during training, post-test phase performance in the quadrant task revealed the consequences of training: perceptual discrimination was reliably enhanced in the Vector group (% change: 12.6; $t(14)=3.05$, $p=0.0087$, $d=0.787$; Fig. 3C), but not in either of the Scalar (% change: 3.17; $t(14)=0.058$, $p=0.954$, $d=0.015$), or Control groups (% change: -0.002; $t(14)=0.058$, $p=0.954$, $d=0.015$). The degree of enhancement was greater for Vector participants than for Control participants ($t(28)=0.027$, $p=0.027$, $d=0.853$; uncorrected), and Scalar and Control did not differ from one another ($t(28)=0.639$, $p=0.528$, $d=0.233$). The difference in improvement did not reach significance for Vector vs. Scalar groups, however ($t(28)=1.66$, $p=0.108$, $d=0.607$). Taken together, these findings suggest that vector error feedback tuned centroid representations in a more general manner, promoting transfer to a separate perceptual task. For completeness, we additionally performed a 2 phase (pre-test, post-test) x 3 condition (Vector, Scalar, Control) ANOVA on RTs, which yielded no significant interaction between phase and condition ($F(2,82)=0.556$, $p=0.576$, $\eta 2=0.013$), although there was a main effect of condition ($F(2,82)=6.14$, $p=0.003$, $\eta 2=0.130$) and marginally significant effect of phase ($F(1,82)=3.84$, $p=0.053$, $\eta 2=0.045$). Follow-up analyses of post-test RTs showed that median response times were lower in the post-test than pre-test for all conditions (ps<0.039) and critically, in the Vector condition ($t14)=3.06$, $p=0.008$, $d=0.791$), arguing against the possibility that a speed-accuracy tradeoff explained the improvement in accuracy for Vector participants. The rate of decrease in RTs (exponential term in model fit) during the training phase did not correlate with the degree of accuracy improvement for Vector participants (r=-0.119, $p=0.673$).

4.2 Experiment 2: Center-of-Mass Discrimination

In Experiment 1, we found that participants who received Vector feedback during training showed better perceptual discrimination of the centroid in the post-test, although this was not accompanied by significant performance changes during training. On the other hand, we observed no learning in either of the Scalar and Control conditions. The current experiment sought to expand these findings in two additional ways: First, to explore the possibility that the lack of improvement during training might reflect ceiling performance in this task, we sought to make the task more difficult. Having found that greater dot location variability was associated with larger errors in Experiment 1, we reasoned that increasing the amount of overall stimulus variability in the display might make the task more difficult, providing a broader dynamic range in which to observe performance changes during the pointing task. We accomplished this by varying the size as well as location of dots, and eliciting estimates of the center-of-mass of the array, where the area of each dot was taken into account. Second, we sought to test the importance of error feedback, per se, to learning. To accomplish this, we included a Guided condition in which participants were passively guided to the correct location of the center-of-mass immediately after array offset. Thus, this condition was identical to the Vector condition except that pointing response error was virtually eliminated by the onset of the response cue.

4.2.1 Methods

Participants

Sixty young adults (40 women, age range: 18-34y) participated.

Stimuli and Procedure

On each trial, participants were presented with an array of 8 dots that varied in both size and location (Fig. 2C). The diameter of individual dots was determined by sampling independently from a Gaussian distribution centered on the base diameter used in Experiment 1 (0.9°) with $\sigma = 0.3°$ (only dots with positively valued diameter were used). Participants were instructed to aim for the center-of-mass of the array, taking into account the area of the dots, and treating the the dots as flat (as opposed to having volume). The center-of-mass is defined as the average of the position coordinates of all dots, weighted by their surface area. The coordinates of the center-of-mass, xc and yc, may be computed as the sum of the product of the individual areas and coordinates of each dot, divided by the total area of all dots on the display:

$$x_c = \frac{\sum m_i x_i}{\sum m_i} \tag{4.4}$$

$$y_c = \frac{\sum m_i y_i}{\sum m_i} \tag{4.5}$$

This experiment included Vector and Scalar conditions (N = 20 each) that were identical to those in Experiment 1. Because the Scalar condition provides a tighter control for the Vector condition than the Control condition (which did not provide trial-by-trial feedback at all), and the two conditions did not differ in Experiment 1, we did not recruit a separate Control group. Another goal of this study was to isolate the contribution of error information in feedback, per se. To accomplish this, we included a Guided condition (N = 20), identical to the Vector condition except that the actual center-of-mass location was revealed

100ms after stimulus offset but prior to response. Guided participants were still instructed to point to the location of the center-of-mass on each trial, but now it was marked with a green crosshair until they made their response. If pairing each dot array with its center-of-mass is sufficient to result in enhanced perceptual discrimination, then Guided participants should show positive transfer to the quadrant task. On the other hand, if response error is necessary to produce transfer, then Guided participants should not show positive transfer to the quadrant task.

Data Analysis

Response data were submitted to the same set of analyses as in Experiment 1. To quantify how much stimulus variability affected center-of-mass estimation in the pointing task during the training phase, here we used multiple regression to model the contribution of horizontal location variability (s_x), vertical location variability (s_y), and variability in dot size (i.e., area, sarea) to error magnitude across trials for each participant ($error = s_{xi}^T \beta_{sx} + s_{yi}^T \beta_{sy} + s_{area_i}^T \beta_{sarea} + \varepsilon_i$).

4.2.2 Results

Pre-test

Pre-test performance on the quadrant task in all groups was significantly better than chance (60.2% correct vs. 25%; $t(59)=32.3$, $p \ll 0.001$, $d=4.17$), and did not differ among groups ($F(2,57)=0.998$, $p=0.375$, $\eta 2=0.034$). Estimates of c also did not differ between groups (Vector: 1.10 cm [s.e.m. = 0.070 cm], Scalar: 1.11 cm [s.e.m. = 0.041 cm]; Guided: 1.20 cm [s.e.m. = 0.094 cm]; $F(2,57)=0.441$, $p=0.645$, $\eta 2=0.015$), nor did median response time ($F(2,57)=2.14$,

p=0.127, η2=0.070).

Training

Overall. During training, participants received either Vector or Scalar error feedback following their response on each trial, or were shown the location of the center-of-mass prior to making their response (Guided). As Guided participants had direct access to the location of the center-of-mass while pointing to the display, their responses provide an approximation of noise in the motor response when perceptual uncertainty about the centroid location is minimized. Overall, Guided participants made substantially smaller pointing errors than Vector and Scalar participants (ts>15.4, p«0.001, d>4.88; Vector RMSE: 1.94 cm; Scalar RMSE: 1.91 cm; Guided RMSE: 0.358 cm). Consistent with Experiment 1, neither overall training phase performance (RMSE; $t(38)$=0.285, p=0.777, d=0.090; Fig. 4A), radial error (Vector: 1.70 cm; Scalar: 1.67 cm; Guided: 0.304 cm; ts<1) nor error variability (Vector: 0.935 cm; Scalar: 0.928 cm; Guided: 0.187 cm; ts<1; Fig. 4B) differed between Vector and Scalar conditions.

Time course. Our overall manipulation of task difficulty was successful (Experiment 2: Vector/Scalar pooled RMSE=1.93 cm vs. Experiment 1: Vector/Scalar/Control pooled RMSE = 1.43 cm; $t(83)$ = 7.23, p<0.001, d=1.57). However, there were no significant performance changes (i.e., nonzero exponential learning rate) over the course of the training phase in either the Scalar ($t(19)$= 1.32, p= 0.203, d=0.295) or Guided conditions ($t(19)$= 0.770, p= 0.451, d=0.172); performance actually tended to get worse in the Vector condition ($t(19)$= 2.40, p= 0.027 d=0.538), perhaps due to fatigue. This shows that increasing stimulus variability was not sufficient to bring about performance changes over the course of the point-

ing task.

Response times significantly decreased over the course of training in the Vector ($t(19)$= 3.77, p=0.001; d=0.843) and Scalar conditions ($t(19)$= 4.95, p<0.001; d=1.11), though there was no significant trend in the Guided condition ($t(19)$= 0.785, p=0.442; d=0.175). This suggests that some component of response planning may generally speed up with additional task practice, but participants likely reached the minimum response time in the Guided condition.

Relationship to stimulus variability. To examine the relationship between stimulus variability and pointing task performance, we again used multiple linear regression to model the contributions of variability in horizontal location, vertical location, and dot area to errors over time. We found that all three of these factors reliably predicted error magnitude for Vector (β_{sx}=0.118: $t(19)$=11.0, p<0.001; β_{sy}=0.150: $t(19)$=16.2, p<0.001 β_{ssize}=0.754; $t(19)$=3.96, p=0.001) and Scalar (βsx=0.112: $t(19)$=14.3, p<0.001; β_{sy}=0.145: $t(19)$=16.1, p<0.001 β_{ssize}=0.719; $t(19)$=3.44, p=0.003) participants, but not Guided (ts<1) participants. To compare the degree to which these factors predicted error magnitude between conditions, we performed a 2 condition (Vector, Scalar) x 3-predictor (horizontal location, vertical location, size) ANOVA, which did not reveal a main effect of condition ($F(1,114)$=0.028, p=0.869, η2=0.0002). This suggests that returning vector error feedback does not necessarily reduce the influence of stimulus variability on error.

Directional bias. Using the Hodges-Ajne test of uniformity in the angular component of errors, we found that neither Vector (p=0.11) nor Scalar groups (p=0.59) showed systematic evidence of directional bias.

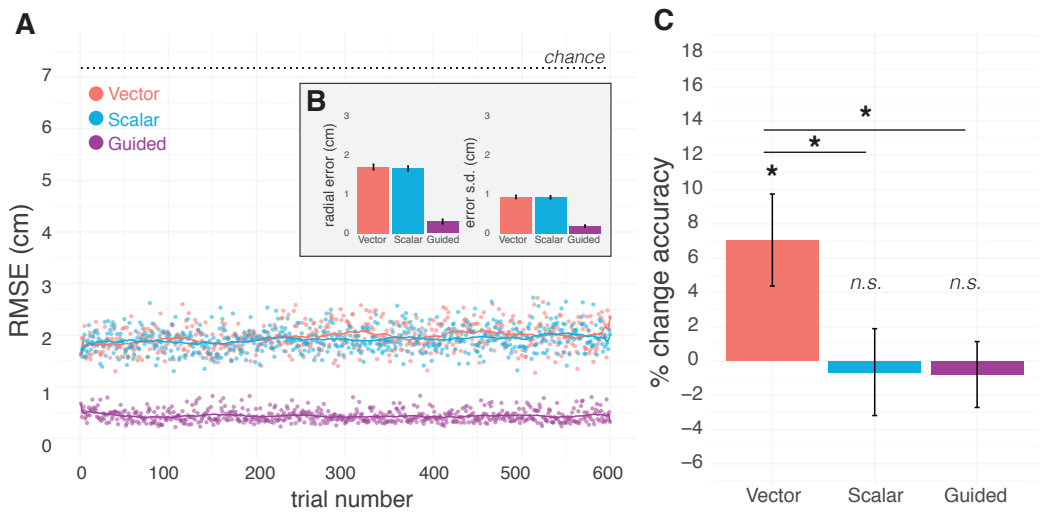


Figure 4.4: Experiment 2 Results. A) Overall performance (RMSE) on the pointing task as a function of trial number for the Vector (Salmon), Scalar (Blue), and Guided (Purple) groups. All groups performed better than chance (dashed line), with Guided participants making substantially smaller errors. B) Performance in the pointing task broken down into accuracy (radial error) and precision (standard deviation). C) Percent change in accuracy between post-test and pre-test in the quadrant task. Error bars represent 1 SEM *<0.05, <0.1.

Post-test

Post-test phase performance was again reliably enhanced in the Vector group (% change: 0.071; $t(19)=2.63$, $p=0.017$, $d=0.588$; Fig. 4C), but not in the Scalar group (% change: -0.007; $t(19)=0.257$, $p=0.800$, $d=0.058$) nor the Guided group (% change: -0.008; $t(14)=0.413$, $p=0.685$, $d=0.092$). The degree of enhancement for Vector participants was marginally greater than for Scalar participants ($t(38)=2.09$, $p=0.044$, $d=0.660$) or Guided participants ($t(38)=2.38$, $p=0.023$, $d=0.752$), and Scalar and Guided performance did not differ from one another ($t(38)=0.045$, $p=0.965$, $d=0.014$). Median response time did not differ between groups ($F(2,57)=0.873$, $p=0.423$, $\eta2=0.030$). These results provide additional evidence that displaying the location of the center-of-mass as vector error feedback during training — beyond merely pairing this location with the dot array or returning scalar error feedback — is necessary to observe positive transfer to another task. We additionally performed a 2 phase (pre-test, post-test) x 3 condition (Vector, Scalar, Guided) ANOVA on RTs, which yielded no significant interaction between phase and condition ($F(2,112)=0.333$, $p=0.717$, $\eta2=0.006$), and no main effects of either condition ($F(2,112)=1.96$, $p=0.145$, $\eta2=0034$) nor phase ($F(1,112)=0.416$, $p=0.520$, $\eta2=0.004$). Median response times were numerically (though not significantly) lower in the post-test than pre-test in the Vector condition ($t(19)=1.19$, $p=0.249$, $d=0.266$), inconsistent with a speed-accuracy tradeoff. The rate of decrease in RTs (exponential term in model fit) during the training phase did not significantly correlate with the degree of accuracy improvement for Vector participants (r=-0.305, $p=0.191$).

4.3 Experiment 3: Role of Pointing Movements in Learning

Experiments 1 and 2 manipulated access to vector error information by returning spatially precise information about the location of the centroid with respect to the location of the participant's response, registered at the point of contact between the participant's fingertip and the display. Although prior studies of statistical summary perception have not involved action toward the summary features, prior work in the field of motor learning suggests that such actions are required to generate sensory-prediction error (for a review, see Shadmehr, Smith, & Krakauer, 2010). Experiment 3 sought to test whether sensory (i.e., haptic and visual) feedback of this kind may have contributed to the learning observed in the first two experiments. Towards this end, we designed a task that elicited perceptual estimates of the centroid during training without a pointing movement towards the display. Instead, participants typed the spatial coordinates of the centroid according to its position in a grid overlaying the display (grid task).

4.3.1 Methods

Participants

Twenty young adults (11 women, age range: 18-34y) participated in Experiment 3.

Stimuli and procedure

The presentation of the dot array was identical to that used in Experiment 1. During the training phase, the entire display was overlaid by a rectangular grid whose cells measured 1.2° x 1.2°, about the size of a fingertip, in order to approximate the precision of responses

afforded by the pointing task. Cells lying within the central 25 x 25-cell region (30.2° x 30.2°) were labeled with an arbitrary 2-character alphanumeric code that varied from trial-to-trial (Fig. 2D). Dot stimuli were of equal size and sampled in an identical manner to that used in Experiment 1. Participants identified the grid cell containing the centroid and typed in the corresponding code. Response entry caused the selected cell to become highlighted with a white border. Error feedback was returned by marking the true location of the centroid with a green crosshair. The grid was visible throughout the entire trial, whereas the 2-character codes were only visible during response. A continuous measure of error was derived by computing the distance between the center of the selected cell and the true centroid on each trial, allowing us to submit these data to the same set of analyses used in Experiments 1 and 2. This metric provides an unbiased measure of response accuracy because the centroid was distributed uniformly across trials, but may result in slight underestimation of response precision owing to additional measurement error. Performance greatly exceeded that expected by chance ($p \ll 0.001$) or by only averaging a subsample of items on the display ($p \ll 0.001$), verifying that participants could perform the task. Prior to beginning the training phase, participants completed 8 practice grid-task trials where only one target dot appeared on each trial and the goal was to select the cell corresponding to this dot's location. The pre-test and post-test phase structure was identical to that used in the first two experiments.

4.3.2 Results

Pre-test

The pre-test and post-test phases were identically structured to those in the first two experiments. Pre-test performance on the quadrant task was significantly better than chance

(60.9% correct vs. 25%; $t(19)=18.0$, $p \ll 0.001$, $d=4.02$), and the value of c (mean = 1.04 cm, s.e.m. = 0.079 cm) did not differ from that of participants in Experiment 1 ($t(63)=0.113$; $p=0.911$; $d=0.03$).

Training

Overall performance (RMSE=1.71 cm; mean radial error=1.47cm; error s.d.=0.862cm; Fig. 5A) greatly exceeded chance ($p \ll 0.001$). Moreover, although the grid task differed from the pointing task both in the presence of gridlines on the display and the type of response, performance was comparable to that for the pointing task in Experiment 1 (Vector RMSE=1.35cm) & Experiment 2 (Vector RMSE=1.94cm). This shows that these changes to the display and response modality did not greatly impact participants' ability to estimate the centroid location. Using the same Hodges-Ajne test of angular uniformity as in the previous experiments, we found no evidence of directional bias ($p=0.591$). As in the previous experiments, there was also no evidence of improvement over the course of training ($t(19)=0.275$, $p=0.787$, $d=0.061$). One potential concern is that if the grid cells were too large, participants might not have treated trials on which they selected the correct grid cell as errors, even though their estimate of the centroid and the true centroid differed. However, the proportion of trials on which the correct grid cell was selected was only 11.1% (s.e.m.=0.63%), showing that the grid cells were small enough for participants to make errors on most trials.

Post-test

If visual comparison (i.e., absent proprioceptive feedback) between the location of the selected grid cell and the precise centroid location provides a sufficiently strong error signal,

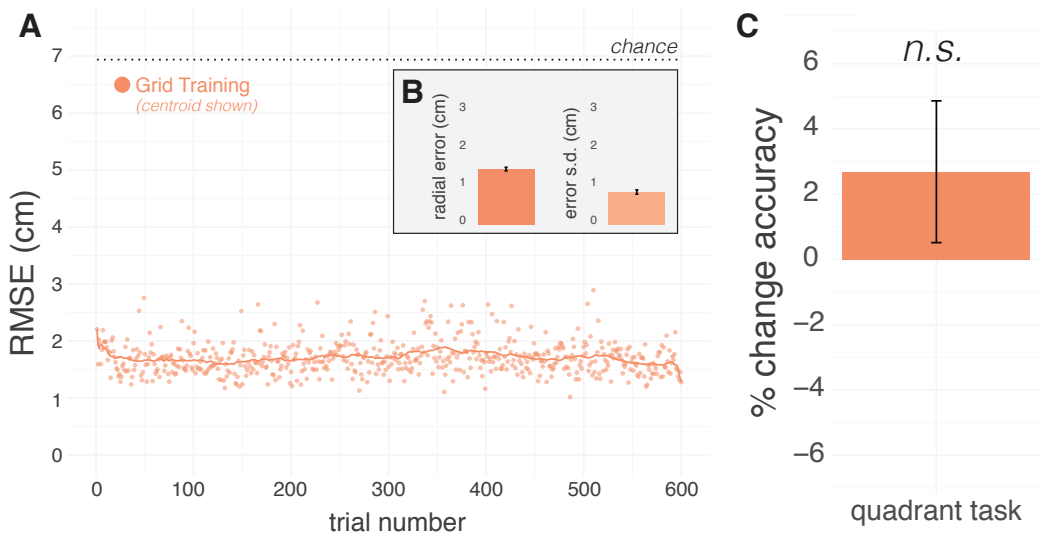


Figure 4.5: Experiment 3 Results. A) Overall performance (RMSE) greatly exceeded chance, where continuous error computed as distance between center of selected grid cell and the centroid. B) Performance in the grid task broken down into accuracy (radial error) and precision (standard deviation). C) Percent change in accuracy between post-test and pre-test in the quadrant task. Error bars represent 1 SEM *<0.05, <0.1.

then training on the grid task should also result in positive transfer to the quadrant task. However, post-test performance did not reliably differ from pre-test performance for participants who trained on the grid task (% change: 2.69; $t(19)=1.24$, $p=0.231$, $d=0.277$; Fig. 5C). These results are consistent with the possibility that sensory-prediction error — conveying information about the mismatch between the expected centroid location (where response movement was aimed) and the actual centroid location (green crosshair) — may play a crucial role in tuning the perceptual representation of the centroid.

4.4 Experiment 4: Change in the Fidelity of Representation

In what ways can training alter centroid representations? Improvements on the quadrant task in the Vector group (Experiments 1 and 2) showed that some change in the perceptual representation occurred, but this was based on a relatively coarse measure — proportion correct in a categorical judgment task. Experiment 4 sought to better quantify changes in the accuracy and precision of the perceptual representation of the centroid. To accomplish this, we used the grid task (from Experiment 3) to collect continuous reports of the centroid location in the pre-test and post-test, allowing us to directly measure the quality of the perceptual representation of the centroid before and after training on the pointing task.

4.4.1 Methods

Participants

Twenty young adults (14 women, age range: 18-34y) participated in Experiment 4.

Stimuli and procedure

Dot stimuli were identical to those used in Experiments 1 and 3. Participants performed the pointing task with vector error feedback during training, and the grid task without feedback during the pre-test and post-test phases (Fig. 2E).

4.4.2 Results

Pre-test

Pre-test performance on the grid task (RMSE=1.92 cm) greatly exceeded chance ($p \ll 0.001$; Fig. 6A) and did not differ from that of participants during the training phase (RMSE=1.71cm) in Experiment 3 ($t(38)=1.38$; $p=0.176$; $d=0.436$).

Training

Likewise, overall pointing task performance (RMSE=1.37cm) during the training phase did not differ from that of the Vector group in Experiment 1 (RMSE=1.43cm; $t(33)=0.156$; $p=0.877$; $d=0.053$). Moreover, radial error (1.20 cm) and error variability (0.659 cm) also did not differ (ts<1). However, participant's pointing errors significantly deviated from angular uniformity in($p=0.03$; 95% CI: [−56.6°, 15.5°]), consistent with an overall bias below and to the right of the centroid. There was no evidence of performance changes (i.e., non-zero learning rate) over the course of the training phase ($t(19)= 1.52$, $p= 0.146$, $d=0.339$).

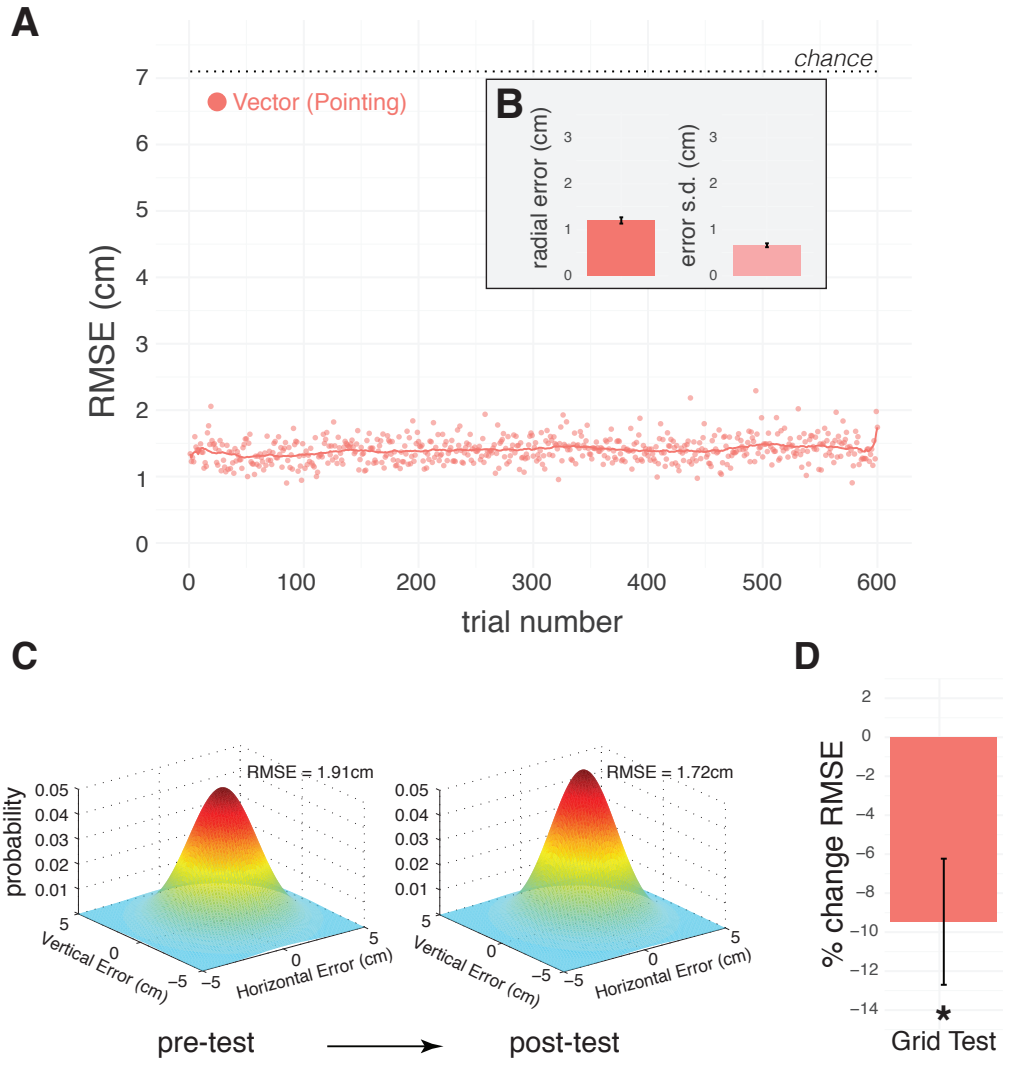


Figure 4.6: Experiment 4 Results. (A) Overall performance (RMSE) greatly exceeded chance on the pointing task with Vector feedback. B) Performance in the pointing task broken down into accuracy (radial error) and precision (standard deviation). (C) Error distribution in the pre-test and post-test modeled as bivariate Gaussian distribution fit to participant errors. (D) Percent change in grid task performance (RMSE) between post-test and pre-test; negative value reflects enhanced fidelity. Error bars represent 1 SEM *<0.05, <0.1.

Post-test

Consistent with the findings in Experiments 1 and 2, training on the pointing task with vector error feedback resulted in generalization to a different task requiring fine perceptual discrimination of the centroid, as quantified by the reduction in RMSE on the grid task in the post-test ($t(19)=2.93$; $p=0.009$; $d=0.654$; Fig. 6C and 5D). This change in overall grid task performance reflected both decreases in average radial error ($t(19)=2.37$; $p=0.029$; $d=0.530$), as well as in the variability of errors ($t(19)=2.80$; $p=0.011$; $d=0.626$).

4.5 General Discussion

In the present study, we sought to understand the role of learning in the formation of statistical summary representations. Specifically, we examined how error-related feedback altered the fidelity of the perceptual representation of the centroid of a set of visual objects. We found converging evidence across four experiments that providing vector error feedback while participants practiced making pointing movements toward the centroid improved the fidelity of their perceptual representation of the centroid, as measured by a separate perceptual test conducted before and after training (quadrant task accuracy in Experiments 1 and 2; grid task RMSE in Experiment 4). By contrast, participants receiving no trial-by-trial feedback (Control condition in Experiment 1) or only scalar feedback (Scalar condition in Experiments 1 and 2) did not show transfer to the separate perceptual test, suggesting that spatially precise error feedback was necessary to improve the fidelity of centroid representations. Critically, feedback had to be in the form of a vector error; participants who were passively guided toward the correct centroid location did not show transfer despite having

precise centroid information (Guided condition in Experiment 2). Moreover, this form of feedback was effective only when coupled with reaching movements towards the visual objects; training on a centroid estimation task that did not involve a pointing movement did not result in learning (Experiment 3). To quantify changes in the perceptual representation, we measured the quality of the perceptual representation before and after training using a continuous measure (Experiment 4) and found that training with vector error feedback enhanced both the accuracy (i.e. reduced bias) and precision (i.e., reduced variability) of centroid estimates. As a whole, these findings suggest that sensory-prediction error — signaling the mismatch between expected and actual consequences of an action — may play a previously unrecognized role in tuning perceptual representations.

4.5.1 Role of feedback in learning

What might underlie the specific advantage for training with vector error feedback? In the Vector condition, the true spatial location of the centroid was precisely marked immediately following a pointing response to the estimated centroid location. When these two points do not coincide, the result is a sensory-prediction error defined as the difference between the pointing movement and the spatial position of the true centroid. On the other hand, in the Scalar condition, numerical feedback was presented that reflected the accuracy of the estimate, but the direction of displacement between the participant's estimate and the true centroid location was not provided. When the expected point score and the actual score earned do not coincide, this type of feedback produces reward-prediction error, reflecting a quantitative difference not tied to a particular sensory modality. Although this type of reward-based feedback has been shown to exert an important influence on learning and

decision-making (Rescorla & Wagner, 1972, Schultz et al., 1997, O'Doherty, 2004) , the role of reward-prediction error in sensorimotor adaptation is less certain (Izawa & Shadmehr, 2011b, Nikooyan & Ahmed, 2015). In the Guided condition, participants were not given the opportunity to actively predict the location of centroid and thus did not experience sensory-prediction error during training. Consistent with the notion that sensory-prediction error may be necessary to induce perceptual learning, Guided participants, despite being provided with exactly the same visual information as Vector participants, did not improve on the perceptual task. An alternative explanation is that Guided participants may have ignored the array, because the task did not demand them to compute its centroid. However, previous investigations of statistical summary perception have shown that observers are able to extract summary features even at short presentation durations (50ms, Chong & Treisman, 2003), leading to the general belief that such features are automatically computed in a pre-attentive manner. If true, participants would not be able to simply disregard this statistical information upon being shown the array (without exerting cognitive control to actively ignore the instructions). Future experiments could directly examine this possibility by including catch trials designed to test whether participants were ignoring the array. Taken together, our findings suggest that the pointing response itself may have been critical for learning, as such visually guided movements provide rich and spatially precise sensory feedback about the position of the fingertip with respect to the centroid target. In the Grid training condition, both haptic feedback about finger position, as well as visual feedback about the location of the response, was less precise than in the Vector condition. Owing to this lack of precise sensory feedback during response, these participants only had access to a crude visual error signal, even though the true location of the centroid was marked on every trial.

These findings are consistent with prior studies of motor adaptation that show that sensory feedback is weighted more heavily when this information is precise (Kording & Wolpert, 2004b, Izawa & Shadmehr, 2011b). One intriguing and counterintuitive aspect of these results is that perceptual performance in the quadrant and grid tasks was enhanced in the Vector condition even though pointing task performance did not significantly improve over the course of the training phase. There are multiple possible reasons for this, though at present we can only offer speculative accounts that could be directly tested in future studies. One straightforward possibility is that a single training phase was insufficient to see significant improvements in response accuracy because such improvement depends on a slow form of visuomotor learning that may be stable within-session but improves over multiple sessions (Karni & Sagi, 1993). A second possibility is that additional engagement of the motor system in the pointing task vs. the grid task may have been required for generalization. Although we have emphasized the informational content carried in error feedback, classic neuropsychological findings (Mishkin et al., 1983, Goodale & Milner, 1992) support a distinction between an action-oriented representation formed along the dorsal visual stream and a more object-based representation along the ventral visual stream. Consistent with this account, it is possible that the 'dorsal representation' supplemented the 'ventral representation' of the centroid during visually-guided reaching to the centroid in the pointing task, yielding high pointing accuracy. Moreover, in the Vector condition, pointing movements in conjunction with precise spatial feedback about the centroid location generated sensory-prediction errors, which may have served to further tune the ventral representation, the benefits of which were only revealed in the perceptual test after training. We also note that while the lack of online improvement during training is puzzling, it was also observed in a prior study

(Berens, 2009) that examined the time course of performance in a task requiring estimation of a statistical summary feature (i.e., mean length of a set of lines) when participants received different kinds of feedback. In that study, some participants were told their response was correct when it was close to the arithmetic mean, while others received feedback based on the quadratic or harmonic means, other mathematically valid ways to compute an average. Participants in the harmonic mean and quadratic mean conditions initially made large errors, which decreased over time. By contrast, participants in the arithmetic mean condition made smaller errors throughout, and did not significantly improve with additional practice. These results suggest that participants default to the arithmetic mean, although binary feedback can nudge them to bias their responses towards a different mean. We use this prior work to support the idea that learning in the Vector condition more likely reflects fine tuning of the centroid representation, rather than coarse calibration to task demands. In the latter case, reduced error would not reflect enhancement in the underlying perceptual representation, but rather participants' figuring out which feature dimensions mattered for task performance. In our study, calibration to the centroid estimation task may have even been accomplished in the first few practice trials prior to the start of training. Nevertheless, vector error feedback may be required for this training to further tune the underlying representation. Given that participants did not improve over the course of 600 trials of training, an open question concerns the amount of training required to give rise to the kind of learning we observed. For example, it is possible that fewer trials may have been sufficient to induce some amount of generalization. On the other hand, if learning driven by sensory-prediction errors proceeds in a more incremental fashion, then a relatively long training session (as long or longer than that used in this study) may be necessary. Future investigations that charac-

terize the 'dose-response' curve relating the amount of training and degree of learning would help to further illuminate the underlying components of error-driven learning in statistical summary perception.

4.5.2 Learning mechanisms

What learning mechanisms might explain how sensory-prediction errors tune the perceptual representation of the centroid? One possibility is that the representation of the individual items in the array is enhanced by directly altering the tuning properties of units residing in early visual areas (Fahle et al., 1995, Gilbert et al., 2001, Karni & Sagi, 1991b). Changes in the representation at that level would be predicted to result in highly specific perceptual expertise at trained locations, but be unlikely to result in generalization to novel locations (e.g., items in the left hemifield, if trained on the right). However, because such early changes alter the fidelity of the inputs to the rest of the system, such training might be predicted to transfer to other tasks involving precise representations of spatial position at these locations (e.g., local change detection). In the present study, we are unable to directly test this possibility, as the locations of individual items varied widely from trial to trial. Nevertheless, testing the spatial specificity of learning in statistical summary perception merits closer examination in future research. Another possibility is that sensory-prediction errors lead to adjustment of the pattern of weights applied to the output from units in early visual areas in accordance with the demands of the task (Dosher & Lu, 2009, Petrov et al., 2005). On this account, errors in the centroid estimation task arise from improper weighting of units coding for specific locations. Most models of learning proceed by incremental updating of the weights associated with active units that contributed to the error (i.e., coding for posi-

tion vectors that are correlated with the error vector), in proportion to error magnitude (i.e., gradient descent). Over several such trials of a centroid estimation task, such error-based updating would result in convergence upon uniform weights, eliminating systematic bias due to improper integration of individual item location information. This kind of learning mechanism would be predicted to facilitate generalization to novel displays with random item locations, but not necessarily to new tasks requiring a different set of weights. Although the centroid may be an already well-learned statistical summary feature for an adult observer, such a learning mechanism may help to explain how novel (Bauer, 2009) or high-level summary features (e.g., mean facial expression, Haberman & Whitney, 2007) are learned. A third possibility is that participants learn statistical properties of the distribution that was most likely to have generated the visual arrays. This form of statistical learning would allow the participant to generate perceptual expectations about the likely location and dispersion of items on each trial, resulting in a more efficient representation of arrays drawn from this distribution (Simoncelli & Olshausen, 2001, Chalk et al., 2010, Michael & de Gardelle, 2014). Such an internal model of stimulus statistics does not prescribe how information about individual item locations should be combined in order to make a response on any given trial. However, it may nevertheless exert an influence on how apparent outliers are weighted in the centroid computation, that is, by discounting these items that appear unlikely to have arisen from the same distribution as the others (Haberman & Whitney, 2010). In this scenario, vector error feedback may serve to correct for such a tendency to discount outliers by increasing the internal estimate of the variability in item locations. This account has intuitive appeal, as it posits that learning occurs by tuning a higher-order statistical representation that mirrors the demands of the task (i.e., estimating summary statistics

from randomly sampled items). However, as an account of error-driven learning, it is complicated by the fact that the correspondence between trial-by-trial vector error feedback and updating of these model parameter estimates is underdetermined. Future studies employing more complex stimulus distributions may elucidate how learning such statistics interacts with feedback. Although we interpret our findings as supporting the notion that spatially precise sensory-error feedback enhanced the underlying centroid representation, these results are also consistent with the possibility that such transfer may only occur in the case of perceptual tasks requiring a discrete response, as in the quadrant task (Experiments 1 and 2) and the grid task (Experiment 4). Indeed, although participants do select a particular grid cell when making their response, their responses carried spatial precision that was on the order of what they were able to select with their fingertip in the pointing task. Thus, the grid task provided quasi-continuous response data that were well suited for measuring continuous changes to the fidelity of the centroid representation, irrespective of the discrete nature of grid cell selection. What kinds of statistical summary representations are most responsive to training? In this study, we focused on how feedback and task practice influenced averaging over a simple visual feature, namely, location. However, it has been previously shown that people are also able to extract averages for complex visual features (Haberman & Whitney, 2007). There is recent evidence that multiple distinct mechanisms might underlie the ability to extract low-level and high-level summary features (Haberman et al., 2015). The relative plasticity of different statistical summary representations may be constrained by the type of mechanism that is responsible for performing these summary operations. In the case of low-level summary perception, these features may be readily and automatically computed (e.g., mean orientation) at a coarse level of resolution, but acuity may be enhanced based on

Table 4.1: Median response time (SEM) by condition and phase, in milliseconds. *Latency to first keystroke in grid task.

		Pre-test	Training	Post-test
E1	Vector	1217 (158)	761 (109)	812 (86)
	Scalar	1031 (105)	771 (54)	843 (99)
	Control	884 (83)	673 (56)	737 (58)
E2	Vector	923 (113)	755 (71)	837 (87)
	Scalar	887 (73)	913 (90)	829 (54)
	Guided	698 (50)	741 (56)	709 (85)
E3	Grid Training	880 (84)	1280 (69)*	796 (67)
E4	Vector	1727 (129)*	852 (70)	1465 (102)*

the actual statistics and range of feature values actually experienced. However, in the case of high-level summary perception, experience and feedback may be especially critical for shaping these summary representations over time. Characterizing the type of feedback and training required to induce improvement in the fidelity of statistical summary representations of abstract, complex visual features would be a fruitful avenue for future research.

4.5.3 Conclusions

Overall, there are two main implications of our findings. First, they suggest that statistical summary representations are not merely an automatic consequence of visual experience, but can be tuned based on external feedback to further optimize performance given specific perceptual goals. Moreover, they show that changes in the fidelity of these perceptual representations may be dependent on the type of sensory error information available during learning, suggesting constraints on the mechanisms by which such tuning may occur.

5

Visual production as a window into perception and learning

When we understand something, we say that we 'see' it. We arrive at the solution to a problem through 'insight.' To better communicate our ideas, we aim to make them 'clear.'

That such metaphors likening cognitive processes to visual experience are so pervasive suggests a close correspondence between how we think about and how we see the world. One of the most compelling demonstrations of this lies in the wide range of technologies humans have devised for creating graphical representations to convey knowledge. Some of these technologies are ancient (e.g., using a stylus to draw or write), while others are quite modern (e.g., computer graphics). The refinement of such representations over many generations has resulted in useful abstractions (e.g., written language, the Cartesian coordinate system) that have enabled sophisticated intellectual activities of broad cultural significance, including: history and law, art and literature, as well as formal mathematics and science (Synofzik et al., 2006, Norman, 1990).

In particular, it may be no accident that major theoretical advances in the sciences have been frequently accompanied by the development of novel visualization tools and visual conventions. For example, when the matrix-style representation of chemical elements in the periodic table was introduced by Mendeleev in 1869, its intuitive layout not only exposed orderly relations among the 66 elements then known, but suggested predictions about the properties of 'missing' elements in the table yet to be discovered (Scerri, 2006).

The goal of this chapter is is to lay out a way to understand and systematically study drawing, the most basic such visualization technique, as a form of visual production that is intimately related to other aspects of cognition, including perception and learning. Drawing has the potential to be illuminating in cognitive psychology for two reasons: (1) drawings are very rich projections from our mental representations, and (2) because drawing generates new representations, generating them may even lead to changes in the mind.

5.1 How do visual recognition and production support communication?

Although the retinal images cast by physical objects and line drawings differ dramatically, humans effortlessly recognize objects in either format. How does the brain accomplish this feat of visual abstraction? Moreover, with just a few well-placed strokes, humans are able to communicate abstract ideas (e.g., object identity) by drawing. What are the mechanisms that underlie the ability to produce a sketch that represents an object?

We examined the possibility that the ability to recognize objects and produce drawings of objects are linked by a common internal substrate — a *generalized object representation*. This premise is plausible, given the reciprocal functions of recognition and production (Fig. 5.1). Specifically, visual recognition entails mapping an image from the external world onto a specific idea in mind; visual production entails expressing a specific idea in mind as an image residing in the external world. We test this hypothesis by evaluating two predictions it makes: (1) that recognition of both objects and human drawings can be achieved by a common visual feature representation; and (2) that training on a novel drawing task can alter this representation, just as training on visual recognition tasks can alter object representations (Goldstone, 1998).

5.2 Experiment 1: Recognizing pictures of objects

People can recognize objects in the face of enormous variation in pose, size, position, lighting, and other sources of noise, a fact which belies the computational difficulty of this feat (Pinto et al., 2008). This ability is supported by a set of hierarchically organized brain regions known as the ventral visual stream (Malach et al., 2002), by which simple visual fea-

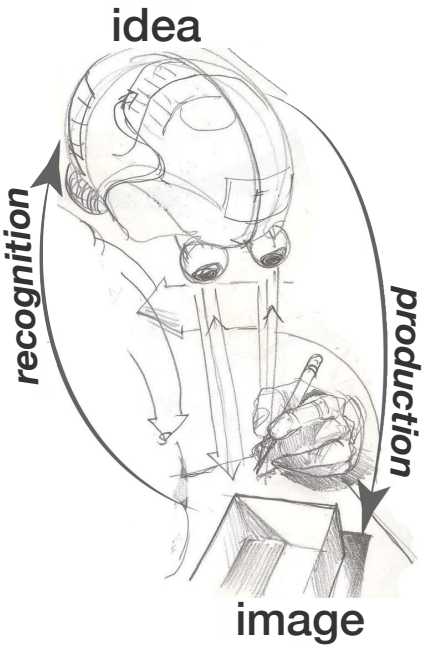


Figure 5.1: Visual recognition entails mapping an image onto a specific idea (e.g., object identity based on a photograph). Visual production entails expressing a specific idea in an image (e.g., identifiable sketch based on an object concept). (Art credit: Jeffery Thompson)

tures (e.g., orientation, spatial frequency) encoded in the lowest area, V1, are successively combined and transformed in such a way as to support read out of abstract object properties (e.g., category, identity) at the top level in the hierarchy, inferior temporal (IT) cortex (Hung et al., 2005).

Supplementing natural visual inputs from objects in the environment, humans have also devised a wide range of technologies for producing pictures that represent objects. The most ancient among these is drawing, whereby lines and marks are made on a surface by manipulating a stylus (Clottes, 2008). Despite large differences between drawings of objects and visual inputs from physical objects (or photorealistic images of objects), they are just as effective at evoking the real-world object (Biederman & Ju, 1988).

What commonalities across line drawings and photorealistic images (e.g., photographs, 3D-renderings) allow people to recognize the same object depicted in such different ways?

Discovering the computational principles that underlie such robust recognition is a challenge that lies at the heart of human visual abstraction. Here we present a computational approach to quantifying such commonalities (Fan et al., 2014, Yamins et al., 2014).

We first assembled a multi-domain imageset containing sketches, photographs, and 3D-rendered images (Fig. 5.2). We then applied a recently developed deep convolutional neural network (Yamins et al., 2014) to extract features from these images (Fig. 5.3a). Inspired by the functional architecture of the ventral visual stream, this model was constructed by efficiently searching the space of hierarchical model architectures for candidates that achieved high performance on challenging object recognition tasks. In addition to achieving human-level performance on these tasks, the higher layers of this model are also quantitatively predictive of neural population responses in high-level visual cortex (e.g., V4 and IT). As such, it was an attractive candidate for investigating the visual invariants that support recognition across image domains.

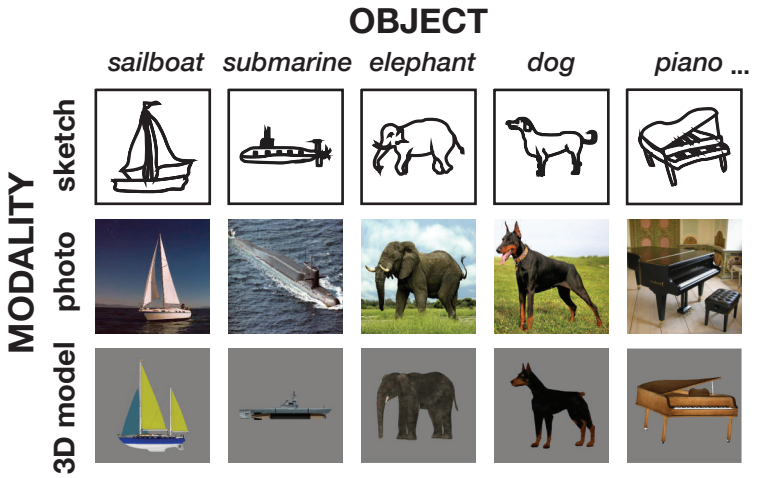


Figure 5.2: Multi-domain imageset containing sketches, photographs, and 3D-rendered images of 105 real-world objects.

Neurally Predictive Model of Object Recognition

We then applied a recently developed deep convolutional neural network model that was inspired by the functional architecture of the ventral visual stream in order to extract features from these images ((Yamins et al., 2014); Fig. 5.3a). This model had been identified using hierarchical modular optimization (HMO), a procedure for efficiently searching among mixtures of convolutional neural networks for candidate hierarchical model architectures that achieve high performance on basic-level object recognition tasks. The HMO procedure was performed on an independent imageset containing photographs only, with no objects in common with the multi-domain imageset described above. In addition to achieving human-level performance on these tasks, the higher layers of the resulting model are also quantitatively predictive of neural population responses in high-level visual cortex (e.g., V4 and IT). As such, it was an attractive candidate for investigating the visual invariants that support recognition across image domains.

5.2.1 Results

The model uses a fixed, but large number of feature dimensions to represent all images. Each image elicits a pattern of feature values at every layer in the model, which may be expressed as a vector in this high-dimensional feature space. For a given image domain, we computed average feature vectors within an object class, then derived correlation matrices based on these feature vectors. This procedure was performed at each of the five layers of the model. Each matrix entry represents the proximity between the average feature vectors from the model for a pair of objects (Kriegeskorte et al., 2008). Higher values (cooler colors) reflect relatively proximal pairs of objects, whereas smaller values reflect more distant object pairs.

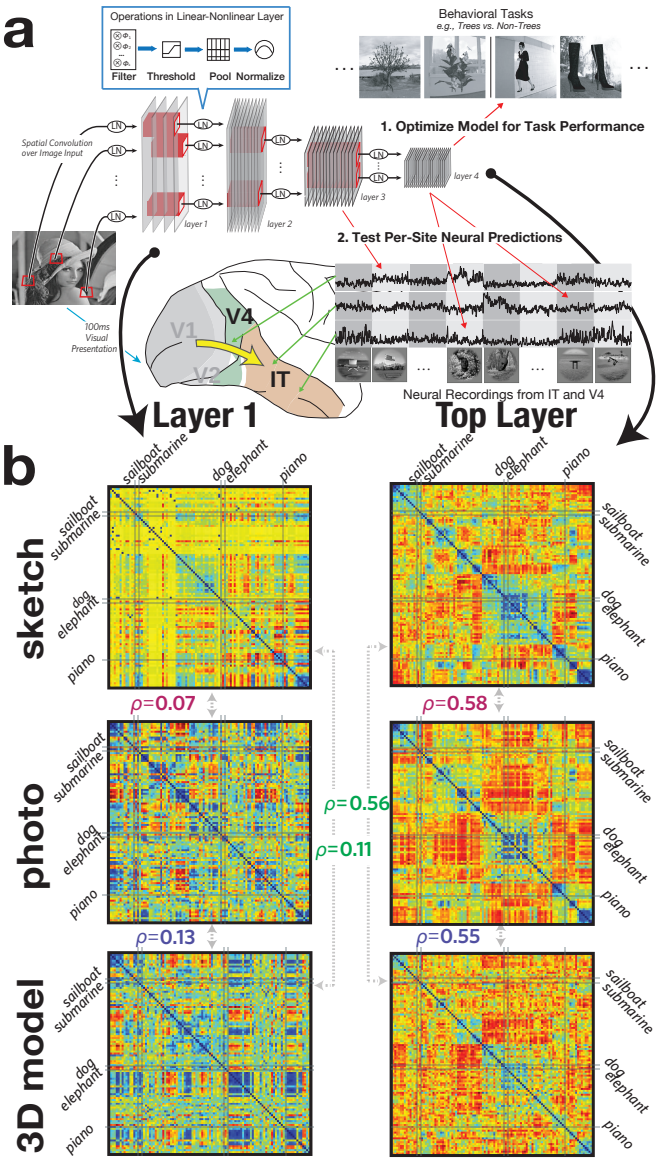


Figure 5.3: (a) Feature extraction using a neurally predictive, deep convolutional neural network model optimized for performance on challenging object recognition tasks. (b) Correlation matrices for each image domain, displaying the overall layout of objects in high-dimensional feature space. Each entry shows correlation distance (1-ρ) between feature vectors for a pair of objects.

Each 105x105 matrix provides a compact visualization of the layout of objects in the high-dimensional feature space inherent to each layer of the model, for each image domain (see Fig. 5.3b for first and top-layer matrices).

All matrices individually show clear block-diagonality, indicating the presence of higher-order structure due to clustering of objects with similar features.

The matrices computed based on top-layer output also show striking cross-domain similarities, both visually and as quantified by correlation distance (Fig. 5.3b, all ps<0.001). This indicates an underlying commonality in the feature representations for the three image modalities at the top layer in the model, the layer whose output has been previously shown to be highly predictive of neural population responses in IT cortex.

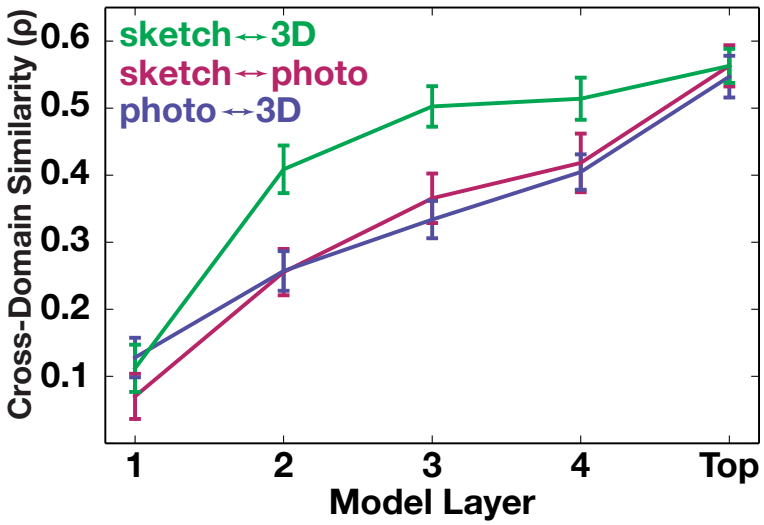


Figure 5.4: Cross-domain similarity between image domains increases as a function of model layer.

By contrast, at the lowest layer in the model, the layer approximating the local/simple features encoded in V1, cross-domain similarities were negligible, though statistically significant $r_{photo,sketch} = 0.086$, p<0.001. This shows that low-level image statistics (e.g., edge

fragments) are insufficient to explain robust recognition across image modalities, especially under conditions of high image variation.

We found that the strength of cross-domain similarities increased over successive layers in the model (Fig. 5.4), quantified as the correlation between layer number and cross-domain similarity for that layer (most critically: $r_{photo,sketch} = 0.895$ $p<0.001$). This is consistent with the understanding that the ventral visual stream computes progressively more abstract properties of objects over successive processing stages. Note that this top-layer cross-similarity is still not at ceiling, however. This indicates that there is still substantial variation in the representation of categories for these image domains. There are multiple possible sources of this variation – it could reflect limitations in the imageset, for instance, the greater category-level variation in the pose, size, and viewpoint in the photo domain relative to the sketches, which tended to be depict objects in a more stereotyped fashion. Additionally, it is not surprising that there is at least some variation across image domains – this would have to be present for viewers to be able to distinguish *between* sketches and photos of objects.

In sum, our results show that a hierarchical neural network model only optimized to recognize photorealistic images of objects generalized to abstract drawings of objects, having produced congruent object-similarity 'maps' across image domains based on the same visual feature representation. These results suggest that the capacity for visual abstraction may be rooted in the functional architecture of the visual system.

5.3 Experiment 2: Producing drawings of objects

How does learning refine object representations? For example, although people tend to label an object at the basic level (Mervis & Rosch, 1981), domain-specific expertise (e.g., knowl-

edge of dogs) makes subordinate-level names (e.g., 'schnauzer') as accessible (if not more accessible) than basic-level names for objects in the domain of expertise (Tanaka & Taylor, 1991). This suggests that initially similar stimuli can become more differentiated with practice. Expertise can also lead to unitization of features that were initially processed separately. For instance, dog experts are worse at recognizing inverted images of dogs than non-experts (Diamond & Carey, 1986), suggesting that extensive experience with an object can lead to automatic binding of features into a viewpoint-specific functional unit.

Such findings suggest that training on recognition tasks can alter object representations. Here we ask whether training on *visual production* tasks can also alter representations of objects. In the previous section, we found that computations approximating those performed by the ventral visual stream produced congruent object similarity 'maps' for both photorealistic images and hand-drawn sketches of objects. Insofar as both the ability to recognize sketches and to produce recognizable sketches recruit a common internal representation, we hypothesized that practice drawing some objects (e.g., horse, cow) might affect the way that those and other, related objects (e.g., sheep) are subsequently represented.

Since recognizability is a key attribute of successful drawings, a natural starting point for examining learning is to identify objects for which untrained participants have trouble producing clearly recognizable drawings — that is, that are frequently confused with drawings of other objects. The most confusable objects are likely to be objects whose drawings share many visual features, even if the objects themselves are not semantically related, *per se* (e.g., bell and pear). To define groups of related objects (i.e., 'visual categories'), we exploit pre-existing object clusters revealed by the model in the original sketch corpus collected by Eitz et al. (2012).

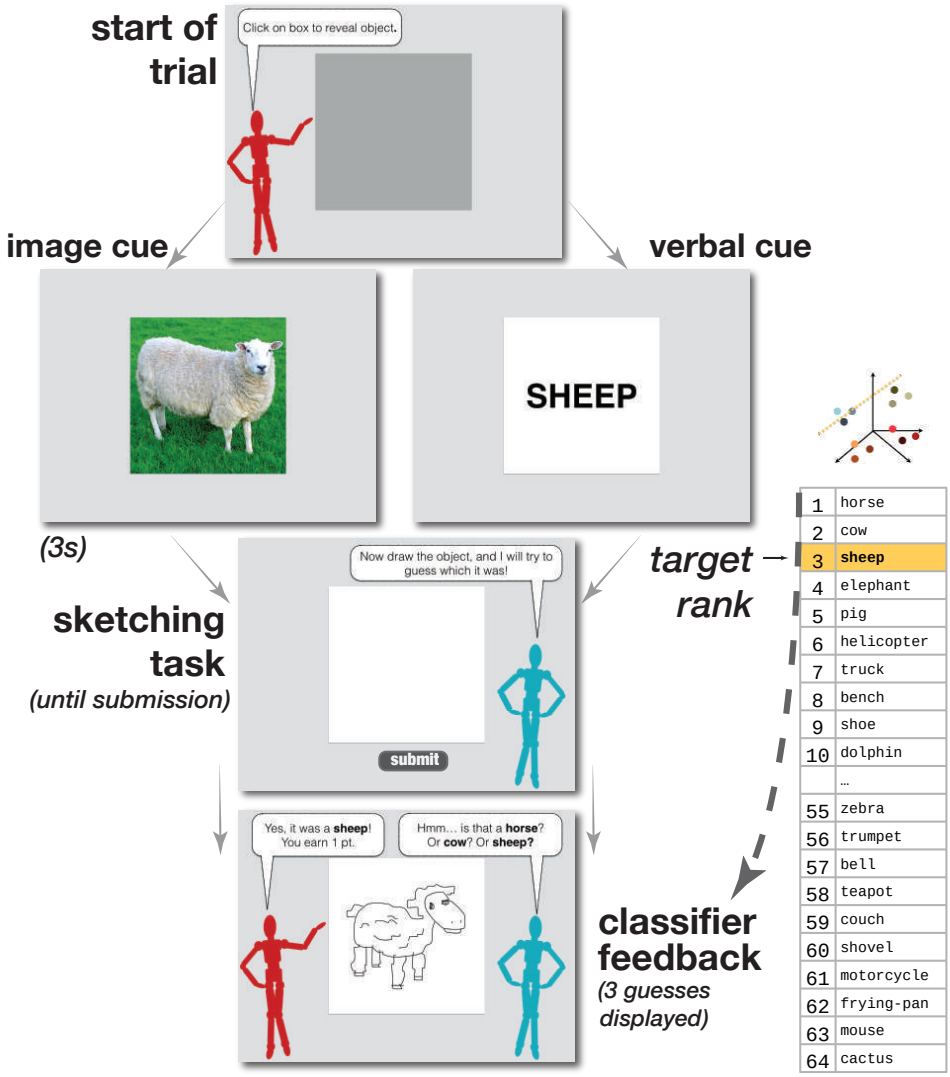


Figure 5.5: Task: On each trial, participants were prompted with a trial-unique photo (N=311) or word (N=282) that referred to a target object for them to sketch. The computer guessed the identity of the drawn object in real time, providing participants immediate feedback about the quality of their sketch. The rank of the target in the list of all 64 guesses (ordered by confidence) returned by the computer was used to track changes in performance across trials.

5.3.1 Methods

Participants

Six hundred and fifty-three participants were recruited via Amazon Mechanical Turk (AMT) for the drawing experiment, with sixty excluded for failing to complete the session. Participants were paid a base amount of $1.50 and up to $3.00 bonus for high task performance. Three hundred and twenty-seven additional participants were recruited (via AMT) to provide labels for the sketches from the drawing experiment, and were paid $0.85 for their participation. All provided informed consent in accordance with the Princeton IRB.

Stimuli and Design

In order to identify groups of objects that are drawn similarly prior to training, we applied a clustering algorithm (affinity propagation with damping=0.9; (Frey & Dueck, 2007)) to the features extracted from the 105-object sketch corpus described above (Eitz et al., 2012). This yielded 16 clusters containing between 3 and 20 objects each. Among clusters containing at least 8 objects, we defined 8 visual categories containing 8 objects each (Fig. 5.6a). Each participant was randomly assigned two of these categories. During training, participants sketched 4 randomly selected objects in one category (Trained) multiple times. Before and after training, participants sketched the other 4 objects in that category (Near), as well as the objects in the second category (Far), once each (Fig. 5.6b).

Task

The sketching task was performed in the context of a game ('Guess My Sketch') in which participants teamed up with two avatars (red, blue) in order to earn points. At the start of each trial, only the red avatar was onscreen. This avatar cued participants with either a trial-unique photo (N=311) or word (N=282) that referred to a target object for them to sketch (Fig. 5.5). After cue offset, the blue avatar appeared, prompting the participant to begin sketching. Upon sketch submission, the blue avatar listed its top three guesses as to the identity of the drawn object, thus providing participants with immediate feedback about the quality of their sketch. These guesses were listed in order of confidence. Participants earned points if any of these guesses were correct, and more points the earlier the correct guess was in the list.

In the image-cue version of the task, unique photographs were used as cues on every trial in order to discourage overly stereotyped sketches. Participants were instructed to "make a sketch in which someone else is likely to recognize the object depicted," but were informed that the sketch did not have to exactly depict what was in the photo. Other than the lack of an image cue, the verbal-cue version of the task was identically structured.

Feedback and Measurement of Drawing Performance

We trained a 64-way support vector machine (SVM) linear classifier on model responses to photographs of the objects used in this study, but no sketches. (Thus, sketch-classification during the experiment reflects pure generalization across image domains.) On each trial, top-layer model features were extracted from the submitted sketch in real time, which were passed to the 64-way classifier to determine feedback. The classifier returned a list of 64 mar-

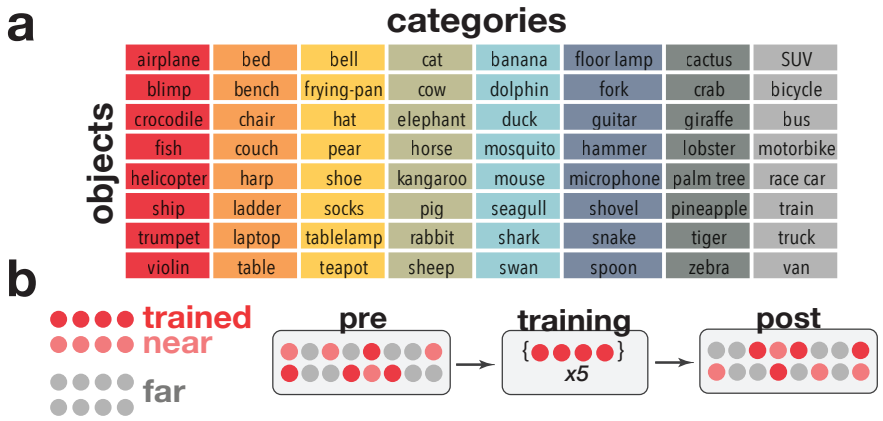


Figure 5.6: (a) Stimuli: Objects belonged to eight visual categories, each containing eight items. (b) Design: Each participant was randomly assigned two of these categories. During training, participants drew four randomly selected objects in one category (Trained) multiple times. Before and after training, participants drew the other four objects in that category (Near), as well as the objects in the second category (Far), once each.

gin values, corresponding to the level of confidence that the test image belonged to each object class. The three objects with the most positive margin values (highest confidence) were returned to the participant as guesses. In the verbal-cue version of the task, when none of the three top guesses were correct, the rank of the target in this ordered object list was also returned to the participant (e.g., *"Too bad... 'giraffe' would have been my 9th guess."*). Because this *target rank* value provides a consistent measure of the 'goodness-of-fit' of the submitted sketch to the target object representation in the model, this value served as our primary measure of task performance. Since the criteria for recognition by the model were fixed, we interpret changes in task performance as reflecting changes to the participants' internal object representations.

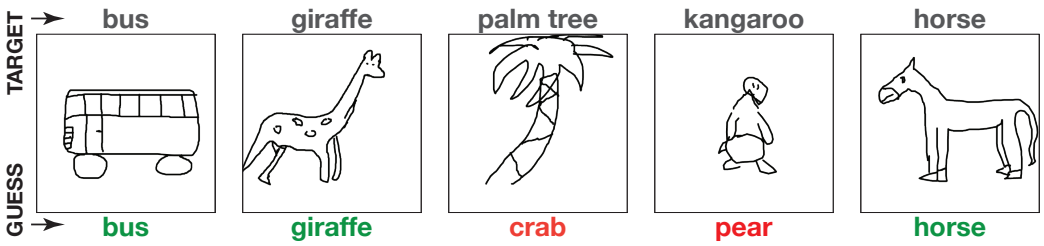


Figure 5.7: Sample sketches from the experiment, with target label and model's top guess.

Validating Model Representation

Because the conditions used by Eitz et al. (2012) to collect sketches differ somewhat from our own (e.g., only verbal cues were used, each participant sketched an object only once, and could apply 'undo', 'redo', 'clear' on their sketches), we first sought to assess the similarity between their sketch corpus and the sketches collected for this experiment. To accomplish this, we extracted features of sketches from the verbal-cue and image-cue versions of the task using top-layer output from the model.

For each version of the task, we computed the average feature vectors for all sketches within an object class, then derived correlation matrices on these feature vectors. In both the image-cue and verbal-cue datasets, we found that sketches of objects within a category were highly similar, validating category assignments. The two matrices were also highly similar to each other (Spearman's ρ=0.890), suggesting that this feature representation successfully captured object identity despite low-level task differences. Moreover, both matrices were highly similar to the original sketch corpus (image-original: ρ=0.715; verbal-original: ρ=0.708). An independent cohort of human participants (N=327) provided three labels to each sketch from the image-cue experiment, in order of confidence, from the set of 64 object labels. We found that human and model recognition performance (d′) was highly consistent

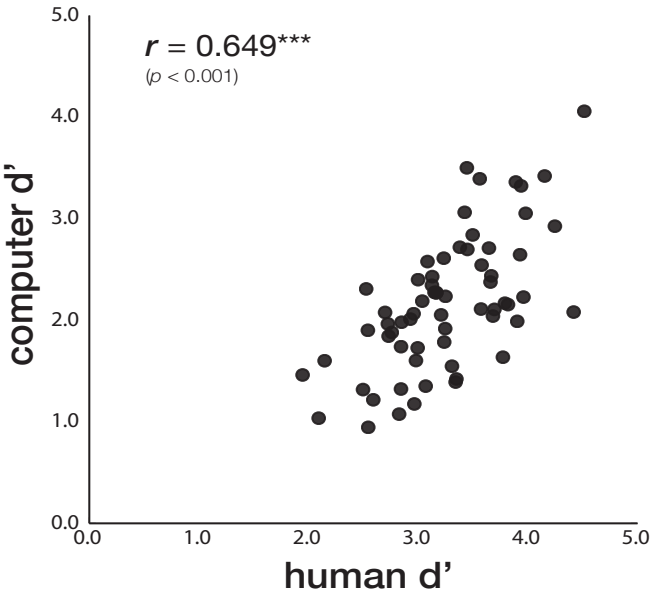


Figure 5.8: An independent cohort of human participants guessed the identity of objects depicted in drawings from the image-cue experiment. Human and computer recognition performance (d') was highly consistent across objects (r=0.649).

across objects (Spearman's ρ=0.649, Fig. 5.8).

Drawing Performance

Since the assignment of objects to condition was randomized across participants, no differences in performance (target rank) on Trained, Near, and Far objects were predicted during the pre-test. To test this, we computed the mean target rank in each condition for each participant (Trained: M=9.66, s.e.m.=0.318; Near: M=9.12, s.e.m.=0.297; Far: M=9.49, s.e.m.=0.254), which we then analyzed using a 3-condition (Trained, Near, Far) x 2 cue-type (image, verbal) repeated-measures ANOVA. There was no main effect of condition on pre-test performance ($F_{2,587}$=1.16, p=0.315). Cue type did have an effect ($F_{2,587}$=14.8, $p<$0.001), with image-cue performance (M=8.69, SD=6.55) exceeding verbal-cue performance (M=10.24, SD=7.58). Cue type did not interact with condition ($F_{2,587}$=0.147, p=0.863), however, so in subsequent analyses we collapsed across cue type. There were no restrictions on how long participants could take to make their drawings, and participants spent on average 30.7s (95% CI: 8.32s,85.6s) to produce each drawing.

Our primary aim in this study was to characterize changes in drawing performance as a function of practice on the Trained objects. Despite the relatively short training period, we hypothesized that participants would learn to produce more recognizable drawings of practiced objects. During the training phase, participants drew Trained objects five times each, in an interleaved order. Otherwise, all aspects of the trial sequence were identical during the training phase as during the pre-test and post-test. In order to test whether performance improved across the five training repetitions of the trained objects, we computed the average rank across objects and participants for each repetition, then estimated the strength of

the directional trend using Spearman's ρ.In order to determine random-effects reliability of this correlation, we used a bootstrapping procedure (Efron & Tibshirani, 1986). On each of 1000 iterations, 593 observers' worth of data were randomly sampled with replacement from the original 593 participants, and Spearman's ρ computed based on this bootstrapped pooled dataset. Across iterations, this procedure yielded an empirical distribution of correlation coefficients. Using this approach, we found that performance improved across the five training repetitions of the trained objects (Spearman's ρ = -0.799; p=0.004).

We additionally sought to understand whether such focused practice would lead to generalized improvement on unpracticed objects. We tested for evidence of generalization at two levels: the *category-level*, as measured by performance changes for Near objects, which belong to the same visual category as Trained objects, and the *task-level*, as measured by Far objects, which belonged to a different category.

To investigate these potential effects, we calculated the change in target rank for each item ($\Delta_{rank} = rank_{post} - rank_{pre}$), then averaged these Δ_{rank} values for each condition within-participant. We then performed the same type of ANOVA as was used for the pre-test analysis, which revealed a highly significant difference among conditions ($F_{2,1182}$=7.67, p<0.001). There was no main effect of cue-type ($F_{1,591}$=1.66, p=0.198), and no interaction with condition ($F_{2,1182}$=0.162, p=0.851), so in subsequent analyses we collapsed across cue-type.

A follow-up t-test revealed that sketches of Trained objects were better recognized by the model following training ($\Delta_{rank} < 0$:t_{592}=4.04, p<0.0001), and this improvement was also statistically reliable when compared with performance on Far objects ($\Delta_{rank,trained} < \Delta_{rank,far}$: t_{592}=2.91, p=0.004). By contrast, model performance for sketches of Near objects did not improve relative to baseline ($\Delta_{rank} < 0$: t_{592}=0.511, p=0.609) and did not differ from

that of Far objects ($\Delta_{rank,near} < \Delta_{rank,far}$: t_{592}=1.15, p=0.252). Recognition of Far objects also did not change significantly relative to baseline ($\Delta_{rank} < 0$: t_{592}=1.22, p=0.223).

Taken together, these results indicate that the training primarily resulted in item-specific benefits, with no clear evidence of generalization to other objects belonging to the same visual category, or generic improvement on the task overall.

If anything, drawings of within-category objects (Near) seemed to worsen relative to objects from a different category (Far). In an exploratory analysis, we restricted our analysis of rank to objects in the same category as the target (resulting in rank scores ranging 1-8, as opposed to 1-64), and found that the decrement in performance for Near objects vs. Far objects was significant ($\Delta_{rank_{category},near} < \Delta_{rank_{category},far}$: t_{592}=2.40, p=0.017). These results are intriguing, providing a preliminary indication that changes in drawing performance may not be a monotonic function of representational distance from trained exemplars.

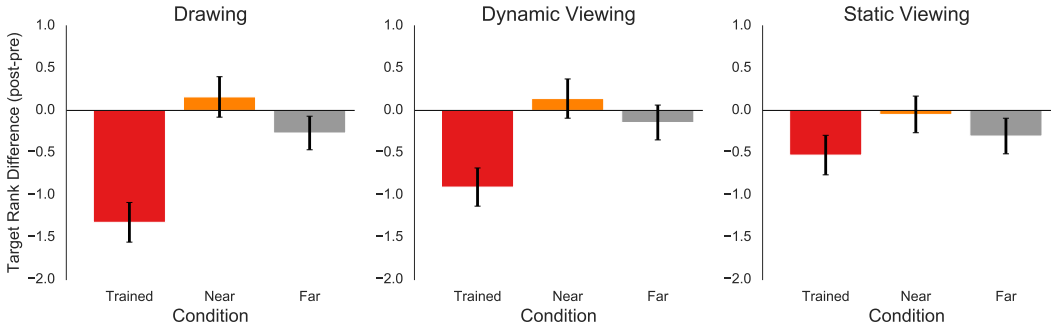


Figure 5.9: Changes in performance between post-test and pre-test for each condition: (*left*) when computing target rank among all 63 distractors and (*right*) when computing target rank relative to remaining 7 objects within the target category only. Error bars represent 1 s.e.m.

Changes to Model Representation

What underlying changes in the feature representation of Trained objects explain improved recognition by the model in the post-test? If we consider each sketch to be a sample from a probability distribution that corresponds to the high-level representation of the target object, there are three intuitive levels at which object representations can change that lead to increased classifier performance for Trained objects: First, the variance of individual Trained object representations might decrease (*within-item*), manifesting as reduced variability across successive sketches of the same object. Second, the centers of Trained object representations might move further apart from one another (*between-item, within-condition*), leading to an increase in the average feature distance among Trained objects, relative to that among Near or Far objects. Third, the Trained objects as a whole might move further from the unpracticed Near objects in the same cluster (*between-condition, within-cluster*), leading to an increase in the average feature distance between Trained and Near objects, relative to corresponding subdivisions of the Far cluster. It should be noted that these sources of representational change need not be mutually exclusive.

To examine each of these possibilities, we first extracted the top-layer feature representation of all sketches using the same model. We measured similarity between any given pair of sketches as the correlation (ρ) between their feature vectors, and derived correlation matrices based on these feature vectors. Conversely, we computed the feature *distance* between any given pair of sketches as 1-(ρ) Synofzik et al. (2006)pKriegeskorte:2008bz.

Within-item. For each Trained object, we constructed the 7x7 feature distance matrix relating successive drawings across the entire session (i.e., 1 pre + 5 training + 1 post), and averaged these feature distance matrices across Trained objects from all participants. We

quantified change in the feature distance between repeated drawings of the same Trained object in two ways: (1) in a discrete fashion, computing the average distance among the 'early' sketches (first three) and the average distance among the 'late' sketches (final three), and taking the difference; as well as (2) in a continuous fashion, using Spearman's rank correlation coefficient, a nonparametric measure of statistical dependence between two variables, to quantify the strength of the negative trend in pairwise feature distances across successive drawings of the same object. A negative trend would indicate 'dampening' in the variability of sketches across repetitions, such that the amount by which each subsequent drawing would differ from the previous decreased over the course of the session. Consistent with the decreasing-variance hypothesis, we found that the feature distance between successive drawings of the same Trained object reliably decreased, measured using both the discrete (Δ=-0.061, p<0.001) and continuous (Spearman's ρ=-0.979, p<0.001) approaches. Moreover, we found that the degree to which drawings became less variable (early vs. late) was significantly correlated with improvement in drawing performance (Δ_{rank}) across participants, ρ=0.187, p<0.001, suggesting that reducing variability in visual production may help mediate object learning.

Between-item, within-condition. For each participant, we first constructed a 32x32 feature similarity matrix relating all 16 object drawings in the pre-test and post-test phases, then averaged the resulting similarity matrix across participants (Fig. 5.11). To evaluate the degree to which Trained object representations had moved further apart from one another, we employed a measure of differentiation that involved anchoring on each Trained object before training and comparing its similarity to other Trained objects before training (pre/pre) vs. after training (pre/post). A decrease in similarity for pre/post vs. pre/pre in the Trained

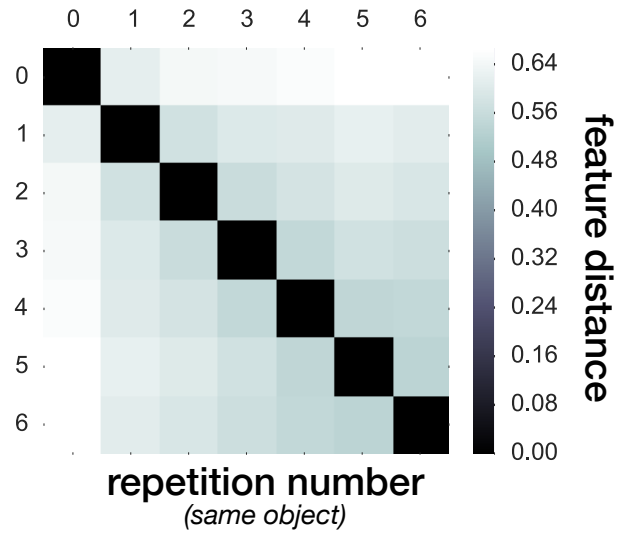


Figure 5.10: Feature distance (1-correlation) between successive drawings of the same object (Trained condition only).

condition that exceeds that for the Near and Far conditions would indicate that drawing practice had differentiated Trained objects from one another. We found that, consistent with this possibility, Trained objects had differentiated to a greater extent than Near objects ($t(592)=3.47$, $p=0.0006$), and more than Far objects ($t(592)=3.40$, $p=0.0007$). This suggests that drawing practice may serve to enhance discriminability between object representations by increasing the distance between practiced objects.

Between-condition, within-cluster. We used a similar approach to evaluate the extent to which Trained objects, *as a whole*, had moved further from the Near objects. To measure differentiation between Trained and Near conditions, we anchored on each Trained and Near object before training and compared its similarity to all other objects in the *other* condition before training (pre/pre) vs. after training (pre/post). To provide a baseline for comparison, we randomly partitioned objects in the Far cluster into two equally-sized sets of

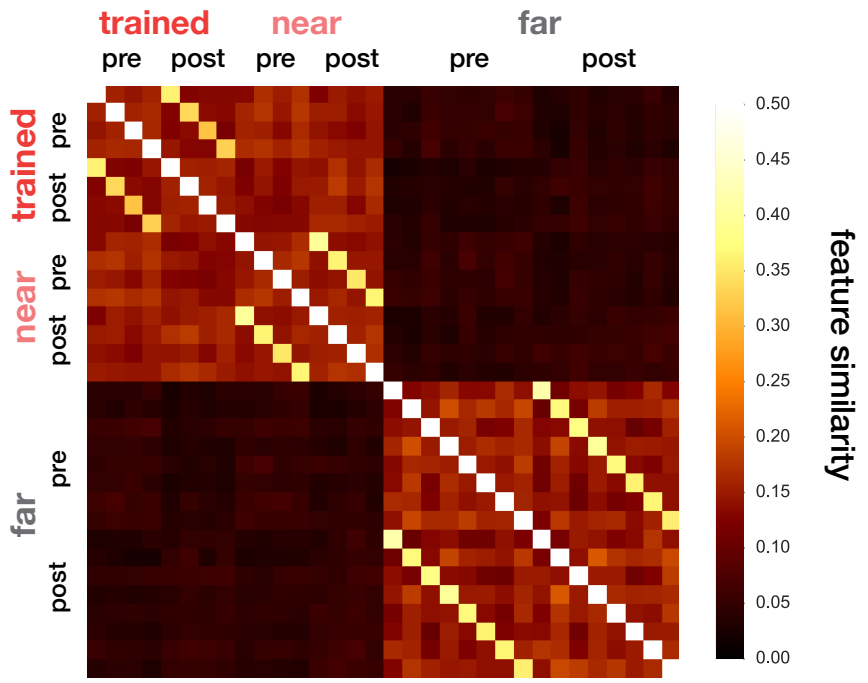


Figure 5.11: Representational similarity between conditions across phases (averaged over object identity): low values in the large off-diagonal blocks correspond to larger average correlation distances between objects in different categories.

objects and evaluated between-'condition', within-cluster differentiation in the same way. A decrease in similarity between Trained and Near relative to within the Far condition would suggest that drawing practice had also moved Trained and Near objects, as a whole, further apart from one another. In support of this possibility, we found that Trained and Near objects had indeed differentiate to a greater degree than had subsets of the Far objects from each other ($t(592)=2.52$, $p=0.012$). This suggests that another factor that underlying improvement in drawing performance may be greater separation between repeatedly drawn objects and other similar, but unpracticed objects.

5.4 Experiment 3: Does visual exposure to sketches suffice to induce differentiation?

What aspects of the drawing task led to improvements in performance? One possibility is that it was not so much the practice drawing, but the exposure to additional sketches of the trained objects that was responsible for these benefits. This would not be so surprising if it turned out to be the case, since most documented forms of perceptual learning are attributable to viewing experience. The current study sought to test this alternative account.

5.4.1 Methods

Participants

Five hundred and ninety-three naive participants were recruited via Amazon Mechanical Turk and compensated in the same manner as in the previous study.

Procedure

For each session in the original cohort, we recruited a new participant to repeat the same sequence of trials, with the only difference being that instead of sketching during the training phase, they would see the final sketch drawn by the participant from the original cohort. To ensure visual processing of these sketches, these participants were instructed to "guess what the computer thinks the sketch looks like" and received bonus points when their guess matched that of the computer. Participants typed their guesses, and only guesses corresponding to one of the 64 valid labels were accepted. Valid labels were displayed in a word bank in the bottom half of the display. The pre-test and post-test phases were identical to those in Experiment 1.

5.4.2 Results

We found that training by viewing the series of sketches produced by the original cohort was insufficient to yield reliable improvement in drawing performance for Trained objects, relative to control Far objects ($t(592)=0.580$, $p=0.562$). Moreover, there was no change in performance for Near objects, relative to control objects ($t(592)=0.714$, $p=0.475$). This shows that visual exposure to sketches was insufficient for improving drawing performance, even when matching the image content and presentation order of sketches. These findings suggest that some dynamic aspect of the drawing task may account for the improvement in the original cohort.

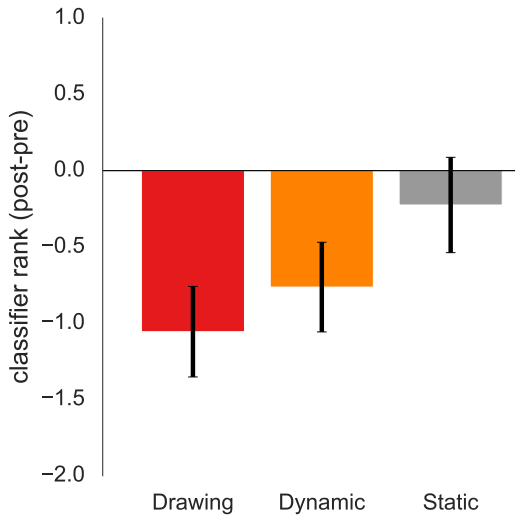


Figure 5.12: Representational similarity between conditions across phases (averaged over object identity): low values in the large off-diagonal blocks correspond to larger average correlation distances between objects in different categories.

5.5 Experiment 4: Does observing sketch production induce learning?

The goal of the current experiment was to estimate the contribution of dynamic visual feedback generated while producing a sketch to improvement in drawing performance. To test this, we recruited a third cohort of naive participants who watched video replays of each drawing being produced, stroke-by-stroke, by their matched participant in the original cohort.

5.5.1 Methods

Participants

Five hundred and ninety-three naive participants were recruited via Amazon Mechanical Turk and compensated in the same manner as in the previous studies.

Procedure

As in the previous study, for each session in the original cohort, we recruited a new participant to repeat the same sequence of trials, with the exception that instead of sketching during the training phase, they would observe a stroke-by-stroke video replay of how the original participant had produced their sketch. Again, these participants were instructed to "guess what the computer thinks the sketch looks like" and received bonus points when their guess matched that of the computer. The pre-test and post-test phases were identical to those in Experiment 1.

5.5.2 Results

These participants did significantly improve on the drawing task, though to a lesser degree than the original cohort ($t(592)=2.09$, $p=0.037$; original Cohen's $d = 0.120$ vs. dynamic observation $d = 0.086$). Again, there was no change in performance for Near objects, relative to control objects ($t(592)=0.751$, $p=0.453$). Taken together with the results of the previous study, these findings suggest that observing the dynamic process of drawing construction plays a specific role in enhancing visual production skill.

5.6 Discussion

The present study sought to understand the relationship between the ability to recognize objects – a biological endowment shared with other species – and to produce images of objects by drawing – a relatively recent technological development from human prehistory. We examined the hypothesis that, despite salient differences between these two functions, that

they are served by a common representational substrate. We tested this by first comparing the high-level feature similarity between photos and drawings of objects using a deep neural network model of ventral visual cortex. The model, while originally only optimized to handle image variation across photos (an approximation to natural variation in visual inputs), produced highly congruent object-category maps for both image domains, suggesting that our capacity to recognize drawings of objects is deeply rooted in the functional architecture of the visual system.

These findings argue against a strong version of the hypothesis that the appearance of line drawings are purely a product of on culturally-specific conventions (Goodman, 1976), even if it is true that there also exist conventions for line drawings. Rather, our results resonate with previous evidence that line drawings can drive visual recognition as strongly as natural visual inputs, including the finding that line drawings are recognized as efficiently as color images of objects (Biederman & Ju, 1988), and that the neural representation of line drawings depicting scenes is similar to that of the corresponding photograph (Walther et al., 2011). The current findings are consistent with prior studies showing that humans from remote communities without cultural exposure to pictorial art (Kennedy & Ross, 1975), infants (Yonas & Arterberry, 1994), and chimpanzees (Tanaka, 2007) can recognize line drawings. At present it seems more likely that the ability to interpret line drawings arises from the same neural mechanisms that evolved to make sense of natural scenes (Sayim, 2011). These include the role of orientation-tuned neurons in extracting edges, accomplished for natural scenes by computing discrete changes in surface attributes (e.g., abrupt transitions from light to dark). Line drawings may capitalize on this pre-existing sensitivity to contours, since these same units also respond similarly to lines (defined by abrupt changes in light-

ness), and even illusory contours (Von der Heydt et al., 1984).

Validation of the model for recognition of drawings equipped us to then use these top-layer features to track changes to participants' representations of objects by measuring feature-level changes to their drawings as a consequence of training. In all experiments, participants performed a visual communication task in which they attempted to convey object concepts to a digital avatar (representing the neural network model) by drawing them. In our first study, training consisted of drawing a subset of these objects repeatedly. We found that practiced objects were better recognized by the model following training, relative to control objects belonging to a different visual category. This improvement was related to greater feature-level differentiation of practiced objects at three levels: variance of individual object representations, distance between practiced objects, and distance between the practiced objects as a whole relative to unpracticed objects from the same visual category.

In order to examine which aspects of the drawing task had led to these improvements, we recruited 593 naïve participants who were each uniquely matched to a participant in the original cohort. One possibility is that repeated visual exposure to drawn images may be sufficient to improve performance. To test this, these new participants repeated the same procedure of drawing before and after training, except that instead of drawing during training, they viewed the final sketch produced by their matched participant. They did not significantly improve, despite receiving identical visual exposure — if not enhanced exposure (they only viewed completed drawings). Another possibility is that the dynamic visual feedback generated while constructing drawings was responsible for the improvement. To test this, we repeated the experiment with a new cohort of 593 participants who watched video replays of each drawing being produced, stroke-by-stroke, by their matched participant. These

participants did significantly improve on the drawing task, though to a lesser degree than the original cohort. Taken together, these additional experiments suggest that experiencing the dynamic process of drawing construction, whether of one's own drawing or via observation of someone else's, may play a specific role in enhancing visual production skill.

These results also suggest that executing the motor movements associated with producing drawings is not necessary to improve (Mattar & Gribble, 2005). However, this might also be attributable to pre-existing familiarity (and prior experience drawing) the real-world objects included in the current study. Future investigations should investigate whether such motor experience may have a stronger effect when learning to draw novel objects. Moreover, the results of the current study suggest that participants do not need to generate the strokes themselves in order to improve, given that observing the dynamic construction of drawings, even if produced by someone else, led to reliable learning. Nevertheless, the current data are not decisive about the role of internally-directed processes, such as memory retrieval for object features and action planning, since these may have nevertheless been engaged during observation if participants were concurrently making predictions about the placement of upcoming strokes, for example. It may be possible that because stimuli were familiar real-world objects, there was already a large degree of overlap in how participants represented them, such that a typical animation of someone else's drawing may have reinforced the same aspects of each object that would have been practiced in the drawing condition. The extent to which generation, as such, may play a significant role in learning (Roediger & Karpicke, 2006b, Slamecka & Graf, 1978), may depend on how novel the objects are (and by extension, how difficult it is to retrieve their component details). To explore the contribution of internally-mediated construction processes on their own, it would be valuable in future

studies to test a condition in which participants are instructed to imagine drawing an object, but not actually produce a drawing.

In the training studies, we did not manipulate the type of feedback returned to participants, which consisted of a ranked list of three object labels. Thus, while differences in feedback cannot explain differences across experiments, it would be valuable in future studies to explore how other forms of feedback may affect learning. For example, if instead of returning a verbal label the avatar were to suggest stroke-level adjustments to the participant's drawing that would improve its score, thereby providing a clearer error signal, this might amplify/accelerate improvement in performance.

A component of visual production not isolated in the current studies includes the motor movements involved in producing a drawing. Our finding that participants were able to improve by observing dynamic replay of another participant's drawing suggests that executing the motor movements is not always necessary for some degree of learning. Nevertheless, to directly estimate the contribution of the motor movements themselves, in a subsequent study we plan to run a variant of the experiment in which participants trace along as each sketch is generated.

Visual production is a complex behavior (Sommers, 1989), recruiting both internally-directed cognitive processes, such as retrieval from existing knowledge, as well as providing a means for externalizing retrieved representations in a richly structured form. The current study understands visual production in these two ways: first, as a route by which people may learn by actively engaging internal representations, and second, as a window into these internal representations.

This first aspect builds on prior work in learning and memory that has demonstrated that

such internally-directed cognitive processes can be beneficial for learning: generating examples from memory tends to promote recall to a greater degree than merely studying the same examples (Crutcher & Healy, 1989, Slamecka & Graf, 1978, Karpicke & Roediger, 2008). Moreover, producing explanations while learning new information has been shown to facilitate reading comprehension (Chi et al., 1994) and causal learning (Legare & Lombrozo, 2014) relative to matched exposure to the same information. Directing internal attention to a feature of a recently-seen object can bias visual long-term memory for that object by suppressing the accessibility of its other features during later recall (Fan & Turk-Browne, 2013a). The current study is also related to prior work on manual exploration of visual objects, which has shown that self-directed sampling of viewing angles enhances later recognition of those objects Harman et al. (1999).

The second aspect constitutes a discovery about the core visual abstractions that are accomplished by the ventral stream, such that even fairly iconic representations of objects, such as the sketches elicited in this and other studies Eitz et al. (2012), can be recognized using a model of object recognition that was only explicitly trained to handle highly photorealistic inputs. Of course, this is not to say that there are no differences between photos and drawings; after all, people can also easily distinguish these types of images. Nevertheless, our findings suggest that one potential reason why even very simple sketches can so effectively evoke real-world objects is that they contain features that coincide with what is computed about objects during natural vision.

Humans draw for many reasons: to depict, to record, to plan, to explain, to create (Tversky, 2010). Drawn images predate the historical record (Clottes, 2008), are pervasive in human culture (Gombrich, 1989), and are often produced prolifically in childhood (Kellogg,

1969). Moreover, drawing is a powerful tool for communication — with just a few strokes it is possible to convey the identity of a face (Bergmann et al., 2013) or express an intention (Galantucci, 2005). Just as investigations of both verbal comprehension and production are indispensable to theories about linguistic communication, a more complete understanding of visual communication will entail examining how visual recognition and production interact to achieve our goals. Ongoing experiments are examining how learning to draw objects affects recognition of those and other objects in demanding recognition tasks, to further evaluate the idea that visual production alters a generalized object representation that supports both recognition and production. Ultimately, inquiries into the psychological basis of visual production may shed new light upon the origins of symbolic writing systems for communication, and the very nature of our ability to apprehend abstract meanings from visual artifacts.

6

Conclusion

The goal of my dissertation research has been to elucidate how the interactive nature of perception, memory, and action supports effective learning. In particular, this work has explored the role of *cognitive actions* — a class of operations performed over mental representations and which have the capacity to transform them through their very execution. As a case study, I focused on object learning — that is, how people acquire knowledge about the

diagnostic features of objects and adaptively exploit this knowledge to achieve their goals. In some cases, these acts are not readily visible by external observers (e.g., memory retrieval). In other cases, these acts are accompanied by motor movements that are externally observable (e.g., reaching) or leave a visible trace (e.g., drawing). Understanding the consequences of such cognitive actions is especially important because they are the essence of intelligent, goal-directed behavior.

Chapters 1-3 described three studies I conducted to examine the consequences of internally-directed cognitive actions (i.e., internal attention) on the organization of long-term memory representations, as well as the subsequent influence of these memories on visual attention and perception. In the first study, we discovered that directing internal attention to one feature of an object biased visual long-term memory (VLTM) for that object by suppressing the accessibility of its other features (e.g., if the location of a red object was practiced, its color was less likely to be remembered later). In the second study, we found that these feature associations in VLTM automatically guided attention during visual search and captured attention even when these features were no longer relevant to the task. In the third study, we found that associations in VLTM and sensory inputs automatically combine to inform perceptual judgments about other, similar objects.

Chapter 4 addressed the influence of task-related and sensory feedback (i.e., visual and haptic) on statistical summary perception. We accomplished this by training participants to estimate the centroid of an array of objects on a touchscreen display under different feedback conditions. Before and after training, they completed a transfer test requiring perceptual discrimination of the centroid. We found that vector error feedback, which conveys error both in terms of distance and direction, was the only form of feedback that improved

perceptual discrimination of the centroid on the transfer test. Moreover, this form of feedback was effective only when coupled with reaching movements toward the visual objects. These findings suggest that sensory-prediction error — signaling the mismatch between expected and actual consequences of an action — may play a previously unrecognized role in tuning perceptual representations.

Chapter 5 reported investigations of how training on *visual production* (i.e., drawing) tasks reshapes the mental representation of visual object categories. We first examined the possibility that a common internal representation of visual objects supported both recognition and production. We found that a model of ventral visual cortex only optimized to recognize objects in photographs generalizes to drawings of objects, suggesting that the capacity for visual abstraction is rooted in the functional architecture of the visual system. Next, we tested the effects of visual production practice and found that repeatedly sketched objects were better recognized after training, and this reflected greater differentiation between them in the model representation.

Traditional studies of learning have examined how people process the information provided to them by the world, yet this leaves aside the many ways in which people actively direct their own learning (Bruner, 1961, Gureckis & Markant, 2012, Papert, 1980). This dissertation provides converging evidence from diverse tasks (i.e., episodic memory, visually-guided pointing, visual production) that cognitive actions, broadly defined, play an important role in determining how well people learn. Moreover, they build on previous work demonstrating the benefits of other kinds of cognitive actions: allocating visual attention to an item during encoding enhances its later recognition (Chun & Turk-Browne, 2007, Uncapher et al., 2011). Retrieving examples from existing knowledge tends to improve recall more

than passively acquiring examples from external sources (Crutcher & Healy, 1989, Slamecka & Graf, 1978, Karpicke & Roediger, 2008). Self-directed exploration of viewing angles onto novel visual objects enhances later recognition of those objects (Harman et al., 1999). Generating explanations while learning new information can improve reading comprehension (Chi et al., 1994) and causal learning (Legare & Lombrozo, 2014) relative to matched exposure to the same information. Goal-directed movements generate sensory feedback, which can directly affect learning and generalization of motor performance to new contexts (Izawa & Shadmehr, 2011b).

However, these actions can also come with a cost: removing attention from an item during encoding not only worsens recognition, but can actually produce a memory cost for this item relative to novel items (Fox, 1995a, Tipper & Cranston, 1985). Recall of one item from memory makes other related items less accessible for later recall, a now-classic effect known as retrieval-induced forgetting (Anderson et al., 1994, 2000). Self-explanations can sometimes impair learning of non-causal properties (Legare & Lombrozo, 2014) and lead to overgeneralization (Williams & Lombrozo, 2013). There are limits to purely exploratory learning: when confronted with difficult categorical distinctions, self-generated examples may not confer any advantage over passively-received ones if learners fail to learn how to select the most diagnostic cases (Markant & Gureckis, 2013).

6.1 From cognitive actions to cognitive technologies

To close, I will briefly describe avenues for future investigation that present opportunities to further generalize our understanding of the integral role cognitive actions play in the mental life of modern humans. I will focus on prospects for a rigorous and quantitative under-

standing of visual production, since this activity is particularly representative of the power and flexibility of cognitive actions, especially when amplified and externalized through technology (Donald, 1991, Kirsh, 2010).

Chapter 5 considered figurative drawing, one of the most ancient forms of visual production. The earliest such images reside in the caves of Sulawesi (Aubert et al., 2014) and Chauvet, and in modern life take the form of drawings, diagrams, maps, and graphs of all varieties, be they digital, printed, or hand-drawn. While preceded by spoken language (Donald, 1991, Pinker, 1995), such primarily graphical modes of visual production predate written language by some 25,000 years. One key feature shared by both drawings and text is that they endure. A spoken word fades as soon as it is uttered. Gestures, which are spontaneously used to augment, modify, or substitute for spoken utterances, are similarly fleeting (Goldin-Meadow & Beilock, 2010). Unlike our private mental images, however, drawings are visible to everyone, and can be shared and further modified by others. Because drawings are necessarily abstractions, across many generations people have collectively figured out how to make drawings and other visualizations to think, communicate, and create — a few of the many ways in which visual production, as a form of *cognitive technology*, helps humans achieve their goals.

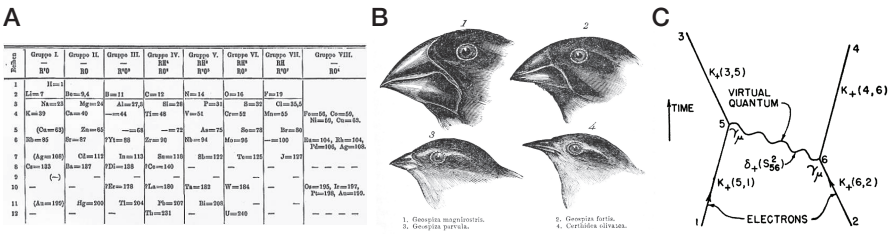


Figure 6.1: A: Mendeleev's Periodic Table as of 1871. B: Drawings of four Galápagos finch specimens highlighting differences in beak morphology. C: A Feynman diagram depicting the exchange of one quantum between two electrons.

6.1.1 Foundation for reasoning and inference

In the history of science, diagrams have played an important role in revealing deeper structure in natural phenomena (Daston & Galison, 2010, Kaiser, 2009). In chemistry, when the matrix-style representation of chemical elements in the periodic table was introduced by Mendeleev in 1869, its intuitive layout not only exposed orderly relations among the 66 elements then known, but suggested properties of 'missing' elements in the table yet to be discovered (Fig. 6.1a, Scerri 2006). In biology, 19th century ornithologist John Gould's now-iconic renderings of the Galápagos finches made differences in beak structure plainly visible, thereby influencing Charles Darwin's thinking about how biological variation and natural selection drive speciation (Fig. 6.1b, Darwin & Jones 2010). In physics, Feynman diagrams provide an intuitive means of simulating the evolution of subatomic-particle states (Fig. 6.1c, Feynman 1949). Since being introduced in the late 1940's, they have become not only adopted widely by practicing theoretical physicists, but also a key pedagogical tool for cultivating the physical intuitions of training physicists (Kaiser, 2009, 2005).

Today, knowing how to make a quantitative graph is a critical component of any practicing scientist's repertoire, as a tool to discover patterns when exploring a complex dataset (Wickham et al., 2010), as well as to communicate empirical findings to other scientists and the public (Johnson et al., 2006). Scientists use graphs in various ways: to query them for numerical information, compare elements, detect clusters, assess relationships between variables, or identify outliers or other anomalies (Springmeyer et al., 1992, Haroz & Whitney, 2012). In many cases, multiple formats may be available for representing the same information. Whereas some formats make interpretation difficult (e.g., numerical tables), others (e.g., bar graphs, scatterplots, heat maps) can expose patterns and nuanced structure (e.g.,

nonlinear relationship, bimodal distribution) that simple summary statistics may obscure (Card et al., 1999). Indeed, figuring out which representations are best suited for expressing the underlying structure in a dataset constitutes a central aspect of scientific discovery.

These historical examples serve as conceptual touchstones for examining the broader role of graphical techniques in human cognition. Future research should address the following questions: First, when trying to understand something, how do people translate their intuitions into the elements of a visualization? Second, how do such visualizations then guide and constrain their understanding?

6.1.2 Graphical commmunication

Diagrams are efficient carriers of meaning — an abstract sketch can effortlessly evoke a real object (Fig. 6.2a), and a simple schematic can convey a complex procedure (Fig. 6.2b). This expressivity is enabled by a shared knowledge about the meaning of graphical forms, such as lines, blobs, and arrows.

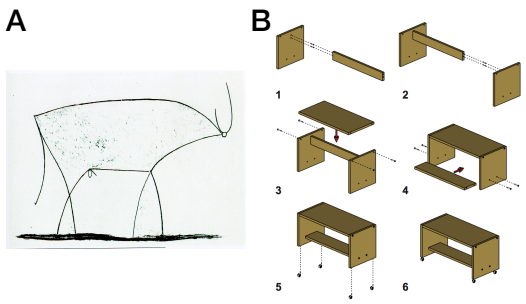


Figure 6.2: A: 'Bull – plate 11' by Pablo Picasso. B: Algorithmically-generated assembly instructions for a TV stand. From Agrawala, M., et al., 2003, ACM Transactions on Graphics, 22(3), 828.

Graphical representations are necessarily selective, emphasizing content most relevant to the topic at hand, while omitting irrelevant details. Of course, which content is considered

to be most relevant may differ depending on the context. For example, a subway map and geographical map prioritize different features of the world, reflecting their different purposes. To help passengers figure out how to get from one station to another, a subway map need only preserve the topological relationships among stations.

Future research should address the following questions: First, how does a pragmatic communicator design her diagrams to be most informative? This is especially important in pedagogical contexts, wherein a teacher who understands what his students know (and do not know) is more effective than one who does not (Shafto et al., 2012). Second, how do interlocutors coordinate shared meanings for graphical forms to achieve a common goal (Wu & Keysar, 2007, Galantucci et al., 2006, Healey et al., 2001)?

6.1.3 Design problem solving

Design sketches are proposals not for what-is, but what-could-be — the architect and engineer's drawing board is the wellspring of the built world (Fig. 6.3). And virtually all designed objects — despite their present materiality — began as ambiguous, untidy, fragmentary sketches full of indeterminate contours, smudges, and blank space. Yet because finished structures do not spring forth whole, their plans must be painstakingly worked out from rough sketch to polished design.

How do exploratory modes of visual production, such as sketching, support the iterative, open-ended, constructive thinking that lies at the heart of design problem-solving? Sketching is a technique that apparently balances two competing objectives — preserving perceptual indeterminacy and progressive constraint satisfaction — a balance that tilts toward the latter as the design process continues. Initially, ambiguity in sketches is considered to be an

asset, as it enables continual reinterpretation and analogical 'leaps' that promote discovery of more original solutions (Tversky et al., 2003, Suwa & Tversky, 2003, Fish & Scrivener, 1990, Goldschmidt, 1992). Meanwhile, each mark set to the page constrains the placement of subsequent marks, and promotes evaluation of the sketch against objective constraints, such as internal consistency and adherence to functional specifications.

These ideas may be formalized in a type of algorithm similar to that used in the field of computer graphics to procedurally generate complex visual content from sets of probabilistic rules (Ritchie et al., 2015a,b). These programs allow designers to rapidly create and explore alternative design possibilities, by iteratively proposing and evaluating candidate solutions on the basis of their estimated performance. Using these programs as a point of departure, future research might construct generative computational models that mirror the dynamics by which humans explore and evaluate potential design solutions.

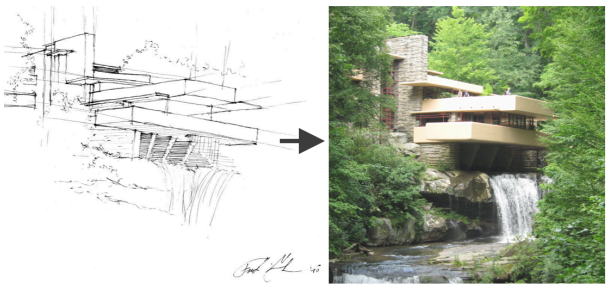


Figure 6.3: Early architectural sketch of "Fallingwater" by Frank Lloyd Wright (left) with recent photograph (right).

Human cognition is inherently dynamic. By investigating the learning principles that govern how people actively make sense of new experiences, we advance our understanding of how the human mind transforms mere information into meaningful knowledge. In the long run, the fruit of these investigations may help to inspire new representational techniques that further expand the intellectual reach of human perception and action.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19, 6.

Al-Aidroos, N., Said, C. P., & Turk-Browne, N. B. (2012). Top-down attention switches coupling between low-level and high-level areas of human visual cortex. *Proceedings of the National Academy of Sciences*, 109(36), 14675–14680.

Alvarez, G. A. & Cavanagh, P. (2004). The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychological Science*, 15(2), 106–111.

Anderson, M. C. (2003). Rethinking interference theory: Executive control and the mechanisms of forgetting. *Journal of Memory and Language*, 49(4), 415–445.

Anderson, M. C., Bjork, E. L., & Bjork, R. A. (2000). Retrieval-induced forgetting: Evidence for a recall-specific mechanism. *Psychonomic Bullein & Review*, 7(3), 522–530.

Anderson, M. C., Bjork, R. A., & Bjork, E. L. (1994). Remembering can cause forgetting: Retrieval dynamics in long-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20(5), 1063–1087.

Anderson, M. C. & Spellman, B. A. (2005). On the status of inhibitory mechanisms in cognition: memory retrieval as a model case. *Psychological Review*, 102(1), 68–100.

Andrews-Hanna, J. R., Reidler, J. S., Huang, C., & Buckner, R. L. (2010). Evidence for the default network's role in spontaneous cognition. *Journal of Neurophysiology*, 104(1), 322–335.

Ariely, D. (2008). Better than average? When can we say that subsampling of items is better than statistical summary representations? *Perception and Psychophysics*, 70(7), 1325–1326.

Aubert, M., Brumm, A., Ramli, M., Sutikna, T., Saptomo, E. W., Hakim, B., Morwood, M. J., van den Bergh, G. D., Kinsley, L., & Dosseto, A. (2014). Pleistocene cave art from Sulawesi, Indonesia. *Nature*, 514(7521), 223–227.

Awh, E. & Jonides, J. (2001). Overlapping mechanisms of attention and spatial working memory. *Trends in Cognitive Sciences*, 5(3), 119–126.

Baddeley, A. D. & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (pp. 47–89). Vol. 8. New York: Academic Press.

Ball, K. & Sekuler, R. (1987). Direction-specific improvement in motion discrimination. *Vision Research*, 27(6), 953–965.

Bauer, B. (2009). The danger of trial-by-trial knowledge of results in perceptual averaging studies. *Attention, Perception, and Psychophysics*, 71(3), 655–665.

Bays, P. M. (2014). Noise in neural populations accounts for errors in working memory. *Journal of Neuroscience*, 34, 10.

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.

Bays, P. M. & Husain, M. (2008). Dynamic shifts of limited working memory resources in human vision. *Science*, 321, 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.

Berens, P. (2009). Circstat: a matlab toolbox for circular statistics. *J Stat Softw*, 31(10), 1–21.

Bergmann, T., Dale, R., & Lupyan, G. (2013). The Impact of Communicative Constraints on the Emergence of a Graphical Communication System. In *Proceedings of the 35th Annual Conference of the Cognitive Science Society* (pp. 1887–1992).

Biederman, I. & Ju, G. (1988). Surface versus edge-based determinants of visual recognition. *Cognitive Psychology*, 20(1), 38–64.

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., Konkle, T., & Alvarez, G. A. (2009). Compression in visual working memory: Using statistical regularities to form more efficient memory representations. *Journal of Experimental Psychology: General*, 138(4), 487–502.

Brady, T. F., Konkle, T., & Alvarez, G. A. (2011). A review of visual memory capacity: Beyond individual items and toward structured representations. *Journal of Vision*, 11(5), 1–34.

Brady, T. F., Konkle, T., Gill, J., Oliva, A., & Alvarez, G. A. (2013). Visual long-term memory has the same limit on fidelity as visual working memory. *Psychological Science*, 24, 6.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, 10, 4.

Brehm, J. W. (1956). Postdecision changes in the desirability of alternatives. *The Journal of Abnormal and Social Psychology*, 52(3), 384–389.

Bruner, J. S. (1961). The act of discovery. *Harvard Educational Review*, 31, 21–32.

Burnham, K. P. & Anderson, D. R. (2004). Multimodel inference: Understanding aic and bic in model selection. *Sociological Methods & Research*, 33, 2.

Card, S. K., Mackinlay, J. D., & Shneiderman, B. (1999). *Readings in Information Visualization: Using Vision to Think*. Morgan Kaufmann.

Carlisle, N. B., Arita, J. T., Pardo, D., & Woodman, G. F. (2011). Attentional Templates in Visual Working Memory. *Journal of Neuroscience*, 31(25), 9315–9322.

Carpenter, S. K., Pashler, H., & Vul, E. (2006). What types of learning are enhanced by a cued recall test? *Psychonomic Bullein & Review*, 13(5), 826–830.

Carrasco, M., Ling, S., & Read, S. (2004). Attention alters appearance. *Nature Neuroscience*, 7(3), 308–313.

Carrier, M. & Pashler, H. (2002). The influence of retrieval on retention. *Memory and Cognition*, 20(6), 633–642.

Cave, C. B. (1997). Very long-lasting priming in picture naming. *Psychological science*, 8(4), 322–325.

Chalk, M., Seitz, A. R., & Series, P. (2010). Rapidly learned stimulus expectations alter perception of motion. *Journal of Vision*, 10, 8.

Chi, M., De Leeuw, N., Chiu, M., & LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science: A Multidisciplinary Journal*, 18(3), 439–477.

Chong, S. C. & Treisman, A. (2003a). Representation of statistical properties. *Vision Research*, 43(4), 393–404.

Chong, S. C. & Treisman, A. (2003b). Representation of statistical properties. *Vision Research*, 43, 4.

Chong, S. C. & Treisman, A. (2005). Statistical processing: computing the average size in perceptual groups. *Vision Research*, 45(7), 891–900.

Christie, J. & Klein, R. (1995). Familiarity and attention: Does what we know affect what we notice? *Memory & Cognition*, 23, 5.

Chun, M. M., Golomb, J. D., & Turk-Browne, N. B. (2011a). A Taxonomy of External and Internal Attention. *Annual Review of Psychology*, 62(1), 73–101.

Chun, M. M., Golomb, J. D., & Turk-Browne, N. B. (2011b). A taxonomy of external and internal attention. *Annual Review of Psychology*, 62(1), 73–101.

Chun, M. M. & Jiang, Y. (1998). Contextual cueing: Implicit learning and memory of visual context guides spatial attention. *Cognitive psychology*, 36(1), 28–71.

Chun, M. M. & Johnson, M. K. (2011). Memory: enduring traces of perceptual and reflective attention. *Neuron*, 72(4), 520–535.

Chun, M. M. & Turk-Browne, N. B. (2007). Interactions between attention and memory. *Current Opinion in Neurobiology*, 17(2), 177–184.

Ciranni, M. A. & Shimamura, A. P. (1999). Retrieval-induced forgetting in episodic memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25(6), 1403–1414.

Clottes, J. (2008). *Cave Art*. Phaidon London.

Crutcher, R. J. & Healy, A. F. (1989). Cognitive operations and the generation effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15(4), 669–675.

Curby, K. M., Glazek, K., & Gauthier, I. (2009). A visual short-term memory advantage for objects of expertise. *Journal of Experimental Psychology: Human Perception and Performance*, 35(1), 94–107.

Darwin, C. & Jones, S. (2010). *The Voyage of the Beagle: Journal of Researches into the Natural History and Geology of the Countries Visited During the Voyage of H.M.S. Beagle Round the World*. Modern Library Classics. Random House Publishing Group.

Daston, L. & Galison, P. (2010). *Objectivity*. Zone Books. Zone Books.

Diamond, R. & Carey, S. (1986). Why faces are and are not special: an effect of expertise. *Journal of Experimental Psychology: General*, 115(2), 107.

Donald, M. (1991). *Origins of the modern mind: Three stages in the evolution of culture and cognition*. Harvard University Press.

Dosher, B. A. & Lu, Z.-l. (2009). Hebbian reweighting on stable representations in perceptual learning. *Learning & Perception*, 1, 1.

Downing, P. E. (2000). Interactions between visual working memory and selective attention. *Psychological Science*, 11(6), 467–473.

Efron, B. & Tibshirani, R. (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical Science*, 54.

Eitz, M., Hays, J., & Alexa, M. (2012). How Do Humans Sketch Objects? *ACM Transactions on Graphics (TOG)*, 31(4), 44.

Endo, N. & Takeda, Y. (2004). Selective learning of spatial configuration and object identity in visual search. *Perception and Psychophysics*, 66(2), 293–302.

Fahle, M., Edelman, S., & Poggio, T. (1995). Fast perceptual learning in hyperacuity. *Vision research*, 35(21), 3003–3013.

Fan, J. E. & Turk-Browne, N. B. (2013a). Internal attention to features in visual short-term memory guides object learning. *Cognition*, 129, 2.

Fan, J. E. & Turk-Browne, N. B. (2013b). Internal attention to features in visual short-term memory guides object learning. *Cognition*, 129(2), 292–308.

Fan, J. E., Yamins, D., DiCarlo, J., & Turk-Browne, N. B. (2014). Mapping core similarity among visual objects across image modalities. In *ACM Siggraph 2014 Posters* (pp.67).: ACM.

Festinger, L. (1957). *A theory of cognitive dissonance*. Evanston, IL: Row & Peterson.

Feynman, R. P. (1949). Space-time approach to quantum electrodynamics. *Physical Review*, 76(6), 769.

Fish, J. & Scrivener, S. (1990). Amplifying the mind's eye: sketching and visual cognition. *Leonardo*, (pp. 117–126).

Fougnie, D., Asplund, C. L., & Marois, R. (2010). What are the units of storage in visual working memory? *Journal of Vision*, 10(12), 27.

Fougnie, D., Suchow, J. W., & Alvarez, G. A. (2012). Variability in the quality of visual working memory. *Nature Communications*, 3, 1229.

Fox, E. (1995a). Negative priming from ignored distractors in visual selection: A review. *Psychonomic Bullein & Review*, 2(2), 145–173.

Fox, E. (1995b). Negative priming from ignored distractors in visual selection: A review. *Psychonomic Bulletin & Review*, 2(2), 145–173.

Frey, B. J. & Dueck, D. (2007). Clustering by passing messages between data points. *science*, 315(5814), 972–976.

Frigge, M., Hoaglin, D. C., & Iglewicz, B. (1989). Some implementations of the boxplot. *American Statistician*, 43(1), 50–54.

Galantucci, B. (2005). An experimental study of the emergence of human communication systems. *Cognitive science*, 29(5), 737–767.

Galantucci, B., Fowler, C. A., & Turvey, M. T. (2006). The motor theory of speech perception reviewed. *Psychonomic Bulletin & Review*, 13(3), 361–377.

Gilbert, C. D., Sigman, M., & Crist, R. E. (2001). The neural basis of perceptual learning. *Neuron*, 31, 5.

Goldin-Meadow, S. & Beilock, S. L. (2010). Action's Influence on Thought: The Case of Gesture. *Perspectives on psychological science*, 5(6), 664–674.

Goldschmidt, G. (1992). Serial sketching: visual problem solving in designing. *Cybernetics and System*, 23(2), 191–219.

Goldstone, R. L. (1998). Perceptual learning. *Annual Review of Psychology*, 49(1), 585–612.

Golomb, J. D., L'Heureux, Z. E., & Kanwisher, N. (2014). Feature-binding errors after eye movements and shifts of attention. *Psychological Science*, 25, 5.

Gombrich, E. H. (1989). *The story of art*. Phaidon Press, Ltd.

Goodale, M. A. & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences*, 15, 1.

Goodman, N. (1976). *Languages of art: An approach to a theory of symbols*. Hackett publishing.

Grier, J. B. (1971). Nonparametric indexes for sensitivity and bias: computing formulas. *Psychological Bulletin*, 75, 424.

Gureckis, T. M. & Markant, D. B. (2012). Self-Directed Learning: A Cognitive and Computational Perspective. *Perspectives on psychological science*, 7(5), 464–481.

Haberman, J., Brady, T. F., & Alvarez, G. A. (2015). Individual differences in ensemble perception reveal multiple, independent levels of ensemble representation. *Journal of Experimental Psychology: General*, 144(2), 432.

Haberman, J. & Whitney, D. (2007). Rapid extraction of mean emotion and gender from sets of faces. *Current Biology*, 17, 17.

Haberman, J. & Whitney, D. (2009). Seeing the mean: Ensemble coding for sets of faces. *Journal of Experimental Psychology: Human Perception and Performance*, 35(3), 718–734.

Haberman, J. & Whitney, D. (2010). The visual system discounts emotional deviants when extracting average expression. *Attention, Perception, & Psychophysics*, 72, 1825–1838.

Hansen, T., Olkkonen, M., Walter, S., & Gegenfurtner, K. R. (2006). Memory modulates color appearance. *Nature Neuroscience*, 9(11), 1367–1368.

Harman, K. L., Keith Humphrey, G., & Goodale, M. A. (1999). Active manual control of object views facilitates visual recognition. *Current Biology*, 9(22), 1315–1318.

Haroz, S. & Whitney, D. (2012). How capacity limits of attention influence information visualization effectiveness. *Visualization and Computer Graphics*.

Healey, P., Swoboda, N., Umata, I., & Katagiri, Y. (2001). Representational form and communicative use. In *Proceedings of the 23rd annual conference of the cognitive science society* (pp. 411–416).: Citeseer.

Held, R. & Gottlieb, N. (1958). Technique for studying adaptation to disarranged hand-eye coordination. *Perceptual and Motor Skills*, 8(3), 83–86.

Herzog, M. H. & Fahle, M. (1997a). The role of feedback in learning a vernier discrimination task. *Vision Research*, 37, 15.

Herzog, M. H. & Fahle, M. (1997b). The role of feedback in learning a vernier discrimination task. *Vision Research*, 37(15), 2133–2141.

Houghton, G. & Tipper, S. P. (1994). : (pp. 53–112). San Diego, CA, US: Academic Press.

Hung, C. P., Kreiman, G., Poggio, T., & DiCarlo, J. J. (2005). Fast readout of object identity from macaque inferior temporal cortex. *Science ...*, 310(5749), 863–866.

Izawa, J. & Shadmehr, R. (2011a). Learning from Sensory and Reward Prediction Errors during Motor Adaptation. *PLoS Computational Biology*, 7(3), e1002012.

Izawa, J. & Shadmehr, R. (2011b). Learning from sensory and reward prediction errors during motor adaptation. *PLoS Computational Biology*, 7, 3.

Jacoby, L. L., Allan, L. G., Collins, J. C., & Larwill, L. K. (1988). Memory influences subjective experience: Noise judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14(2), 240.

Johnson, C., Moorhead, R., Munzner, T., Pfister, H., Rheingans, P., & Yoo, T. S. (2006). NIH/NSF Visualization Research Challenges Report.

Johnson, M. K., Reeder, J. A., Raye, C. L., & Mitchell, K. J. (2002). Second thoughts versus second looks: An age-related deficit in reflectively refreshing just-active information. *Psychological Science*, 13, 63–66.

Kaiser, D. (2005). Making tools travel: Pedagogy and the transfer of skills in postwar theoretical physics. *Pedagogy and the Practice of Science: Historical and Contemporary Perspectives*, (pp. 41–74).

Kaiser, D. (2009). *Drawing theories apart: the dispersion of Feynman diagrams in postwar physics*. University of Chicago Press.

Karni, A. & Sagi, D. (1991a). Where practice makes perfect in texture discrimination: evidence for primary visual cortex plasticity. *Proceedings of the National Academy of Sciences*, 88(11), 4966–4970.

Karni, A. & Sagi, D. (1991b). Where practice makes perfect in texture discrimination: evidence for primary visual cortex plasticity. In *4970* 88(11), 4966: Proceedings of the National Academy of Sciences of the United States of America.

Karni, A. & Sagi, D. (1993). The time course of learning a visual skill. *Nature*, 365, 6443.

Karpicke, J. & Roediger, H. (2008). The critical importance of retrieval for learning. *Science*, 319, 966–968.

Kellogg, R. (1969). *Analyzing children's art*. National Press Books Palo Alto, CA.

Kennedy, J. M. & Ross, A. S. (1975). Outline picture perception by the songe of papua. *Perception*, 4(4), 391–406.

Kirsh, D. (2010). Thinking with external representations. *AI & SOCIETY*, 25(4), 441–454.

Klinger, E. & Cox, W. M. (1987). Dimensions of thought flow in everyday life. *Imagination, Cognition and Personality*, 7(2), 105–128.

Knill, D. C. & Pouget, A. (2004). The Bayesian brain: the role of uncertainty in neural coding and computation. *Trends in Neurosciences*, 27(12), 1552–1566.

Kording, K. P. & Wolpert, D. M. (2004a). Bayesian integration in sensorimotor learning. *Nature*, 427(6971), 244–247.

Kording, K. P. & Wolpert, D. M. (2004b). Bayesian integration in sensorimotor learning. *Nature*, 427, 6971.

Kriegeskorte, N., Mur, M., Ruff, D. A., Kiani, R., Bodurka, J., Esteky, H., Tanaka, K., & Bandettini, P. A. (2008). Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron*, 60(6), 1126–1141.

Lavie, N. & Fox, E. (2000). The role of perceptual load in negative priming. *Journal of Experimental Psychology: Human Perception and Performance*, 26(3), 1038.

Legare, C. H. & Lombrozo, T. (2014). Journal of Experimental Child Psychology. *Journal of Experimental Child Psychology*, 126(C), 198–212.

Lepsien, J. & Nobre, A. C. (2006). Attentional modulation of object representations in working memory. *Cerebral Cortex*, 17(9), 2072–2083.

Levy, B. J. & Anderson, M. C. (2002). Inhibitory processes and the control of memory retrieval. *Trends in Cognitive Sciences*, 6(7), 299–305.

Lieberman, M. D., Ochsner, K. N., Gilbert, D. T., & Schacter, D. L. (2001). Do amnesics exhibit cognitive dissonance reduction? *The role of explicit memory and attention in attitude change. Psychological Science*, 12(2), 135–140.

Liu, J., Dosher, B., & Lu, Z.-L. (2014). Modeling trial by trial and block feedback in perceptual learning. *Vision Research*, 99(C), 46–56.

Luck, S. J. & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, 390, 279–280.

MacLeod, C. M. (1998). Directed forgetting. In J. M. Golding & C. M. MacLeod (Eds.), *Intentional forgetting: Interdisciplinary approaches*. Mahwah, NJ: Lawrence Erlbaum Associates.

Malach, R., Levy, I., & Hasson, U. (2002). The topography of high-order human object areas. *Trends in Cognitive Sciences*, 6(4), 176–184.

Markant, D. B. & Gureckis, T. M. (2013). Is It Better to Select or to Receive? Learning via Active and Passive Hypothesis Testing. *Journal of experimental psychology. General*.

Mason, M. F., Norton, M. I., Van Horn, J. D., Wegner, D. M., Grafton, S. T., & Macrae, C. N. (2007). Wandering minds: The default network and stimulus-independent thought. *Science*, 315, 393–395.

Mattar, A. A. G. & Gribble, P. L. (2005). Motor Learning by Observing. *Neuron*, 46(1), 153–160.

Mervis, C. B. & Rosch, E. (1981). Categorization of natural objects. *Annual review of Psychology*, 32(1), 89–115.

Michael, E. & de Gardelle, V. (2014). Priming by the variability of visual information. In *7878* 111(21), 7873: Proceedings of the National Academy of Sciences of the United States of America.

Mishkin, M., Ungerleider, L., & Macko, K. (1983). Object vision and spatial vision: Two cortical pathways. *Trends in Neurosciences*, (pp. 414–417).

Moores, E., Laiti, L., & Chelazzi, L. (2003). Associative knowledge controls deployment of visual selective attention. *Nature Neuroscience*, 6, 2.

Most, S. B., Scholl, B. J., Clifford, E. R., & Simons, D. J. (2005). What you see is what you set: sustained inattentional blindness and the capture of awareness. *Psychological Review*, 112(1), 217–242.

Myczek, K. & Simons, D. J. (2008). Better than average: Alternatives to statistical summary representations for rapid judgments of average size. *Perception and Psychophysics*, 70(5), 772–788.

Neill, W. T. & Valdes, L. A. (1992). Persistence of negative priming: Steady state or decay? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18(3), 565–576.

Nikooyan, A. A. & Ahmed, A. A. (2015). Reward feedback accelerates motor learning. *Journal of Neurophysiology*, 113(2), 633–646.

Nobre, A., Coull, J., Maquet, P., Frith, C., Vandenberghe, R., & Mesulam, M. (2004a). Orienting attention to locations in perceptual versus mental representations. *Journal of cognitive neuroscience*, 16(3), 363–373.

Nobre, A. C., Coull, J., Maquet, P., Frith, C., Vandenberghe, R., & Mesulam, M. (2004b). Orienting attention to locations in perceptual versus mental representations. *Journal of Cognitive Neuroscience*, 16(3), 363–373.

Norman, D. A. (1990). *Cognitive artifacts*. Department of Cognitive Science, University of California, San Diego.

Norman, K. & Newman, E. (2007). A neural network model of retrieval-induced forgetting. *Psychological Review*, 114(4), 887–953.

Nungester, R. J. & Duchastel, P. C. (1982). Testing versus review: Effects on retention. *Journal of Educational Psychology*, 74(1), 18–22.

O'Doherty, J. P. (2004). Reward representations and reward-related learning in the human brain: insights from neuroimaging. *Current Opinion in Neurobiology*, 14, 6.

Olivers, C. N. (2011). Long-term visual associations affect attentional guidance. *Acta Psychologica*, 137, 2.

Olivers, C. N. L., Meijer, F., & Theeuwes, J. (2006). Feature-based memory-driven attentional capture: Visual working memory content affects visual attention. *Journal of Experimental Psychology: Human Perception and Performance*, 32(5), 1243–1265.

Olson, I. R. & Jiang, Y. (2002). Is visual short-term memory object based? *Rejection of the -strong-object" hypothesis. Attention, Perception, and Psychophysics*, 64(7), 1055–1067.

Papert, S. (1980). *Mindstorms: Children, computers, and powerful ideas*. Basic Books, Inc.

Parkes, L., Lund, J., Angelucci, A., Solomon, J. A., & Morgan, M. (2001). Compulsory averaging of crowded orientation signals in human vision. *Nature Neuroscience*, 4(7), 739–744.

Pelli, D. G. (1997). The videotoolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, 10, 4.

Petrov, A. A., Dosher, B. A., & Lu, Z.-l. (2005). The dynamics of perceptual learning: an incremental reweighting model. *Psychological Review*, 112, 4.

Pinker, S. (1995). *The language instinct: The new science of language and mind*, volume 7529. Penguin UK.

Pinto, N., Cox, D. D., & DiCarlo, J. J. (2008). Why is real-world visual object recognition hard? *PLoS Computational Biology*, 4(1), e27.

Posner, M., Snyder, C., & Davidson, B. (1980). Attention and the detection of signals. *Journal of experimental psychology. General*, 109(2), 160.

Rappaport, S. J., Humphreys, G. W., & Riddoch, M. J. (2013). The attraction of yellow corn: Reduced attentional constraints on coding learned conjunctive relations. *Journal of Experimental Psychology: Human Perception and Performance*, 39, 4.

Rescorla, R. A. & Wagner, A. R. (1972). A theory of pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement.Â classical conditioning ii: Current research and theory. *Â*, 2, 64–99.

Rinaman, W., Heil, C., Strauss, M., Mascagni, M., & Sousa, M. (2012). *Probability and Statistics*. In CRC Standard Mathematical Tables and Formulae. CRC Press, 30th edition.

Ritchie, D., Lin, S., & Goodman, N. (2015a). Generating Design Suggestions under Tight Constraints with Gradient-based Probabilistic Programming. *Computer Graphics ....*

Ritchie, D., Mildenhall, B., Goodman, N., & Hanrahan, P. (2015b). Controlling Procedural Modeling Programs with Stochastically-Ordered Sequential Monte Carlo. (pp. 1–11).

Roediger, H. L. & Karpicke, J. D. (2006a). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science*, 1(3), 181–210.

Roediger, H. L. & Karpicke, J. D. (2006b). Test-enhanced learning taking memory tests improves long-term retention. *Psychological science*, 17(3), 249–255.

Sagi, D. (2011). Perceptual learning in vision research. *Vision Research*, 51(13), 1552–1566.

Sahakyan, L., Waldum, E. R., Benjamin, A. S., & Bickett, S. P. (2009). Where is the forgetting with list-method directed forgetting in recognition? *Memory and Cognition*, 37(4), 464–476.

Saslow, M. G. (1967). Latency for saccadic eye movement. *Journal of the Optical Society of America*, 57, 8.

Sayim, B. (2011). What line drawings reveal about the visual brain. (pp. 1–4).

Scerri, E. R. (2006). *The periodic table: its story and its significance*. Oxford University Press.

Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science*, 275, 5306.

Seitz, A. R., Nanez, Sr, J. E., Holloway, S., Tsushima, Y., & Watanabe, T. (2006). Two cases requiring external reinforcement in perceptual learning. *Journal of Vision*, 6(9), 966–973.

Shadmehr, R., Smith, M. A., & Krakauer, J. W. (2010). Error correction, sensory prediction, and adaptation in motor control. *Annual review of neuroscience*, 33, 89–108.

Shafto, P., Goodman, N., & Frank, M. (2012). Learning From Others: The Consequences of Psychological Reasoning for Human Learning. *Perspectives on psychological science*, 7(4), 341–351.

Shiffrin, R. M. & Atkinson, R. (1969a). Storage and retrieval processes in long-term memory. *Psychological Review*, 76(2), 179.

Shiffrin, R. M. & Atkinson, R. C. (1969b). Storage and retrieval processes in long-term memory. *Psychological Review*, 76(2), 179–193.

Shiu, L. P. & Pashler, H. (1992). Improvement in line orientation discrimination is retinally local but dependent on cognitive set. *Perception and Psychophysics*, 52(5), 582–588.

Simoncelli, E. P. & Olshausen, B. A. (2001). Natural image statistics and neural representation. *Annual Review of Neuroscience*, 24, 1.

Singer, J. L. (1966). *Daydreaming: An introduction to the experimental study of inner experience*. New York, NY: Crown.

Slamecka, N. J. & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning and Memory*, 4(6), 592–604.

Sommers, P. v. (1989). A system for drawing and drawing-related neuropsychology. *Cognitive Neuropsychology*, 6(2), 117–164.

Soto, D., Hodsoll, J., Rotshtein, P., & Humphreys, G. W. (2008a). Automatic guidance of attention from working memory. *Trends in Cognitive Sciences*, 12(9), 342–348.

Soto, D., Hodsoll, J., Rotshtein, P., & Humphreys, G. W. (2008b). Automatic guidance of attention from working memory. *Trends in Cognitive Sciences*, 12(9), 342–348.

Soto, D., Humphreys, G., & Rotshtein, P. (2007). Dissociating the neural mechanisms of memory-based guidance of visual selection. *Proceedings of the National Academy of Sciences*, 104(43), 17186–17191.

Soto, D. & Humphreys, G. W. (2007). Automatic guidance of visual attention from verbal working memory. *Journal of Experimental Psychology: Human Perception and Performance*, 33, 3.

Springmeyer, R. R., Blattner, M. M., & Max, N. L. (1992). A characterization of the scientific data analysis process. In *Proceedings of the 3rd conference on Visualization'92* (pp. 235–242).: IEEE Computer Society Press.

Stokes, M. G., Atherton, K., Patai, E. Z., & Nobre, A. C. (2012). Long-term memory prepares neural activity for perception. In *E367* 109(6), E360: Proceedings of the National Academy of Sciences.

Strube, M. J. & Bobko, P. (1989). Testing hypotheses about ordinal interactions: Simulations and further comments. *Journal of Applied Psychology*, 74, 2.

Suchow, J. W., Brady, T. F., Fougnie, D., & Alvarez, G. A. (2013). Modeling visual working memory with the memtoolbox. *Journal of Vision*, 13(10), 1–8.

Suchow, J. W., Fougnie, D., & Alvarez, G. A. (2012). Visual working metamemory. *Journal of Vision*, 12(9), 348–348.

Suengas, A. & Johnson, M. (1988). Qualitative effects of rehearsal on memories for perceived and imagined complex events. *Journal of experimental psychology. General*, 117(4), 377–389.

Summerfield, J. J., Lepsien, J., Gitelman, D. R., Mesulam, M. M., & Nobre, A. C. (2006). Orienting attention based on long-term memory experience. *Neuron*, 49(6), 905–916.

Suwa, M. & Tversky, B. (2003). Constructive perception: A metacognitive skill for coordinating perception and conception. In *Proceedings of the Annual Conference of the Cognitive Science Society, Boston*.

Synofzik, M., Thier, P., & Lindner, A. (2006). Internalizing agency of self-action: perception of one's own hand movements depends on an adaptable prediction about the sensory action outcome. *Journal of neurophysiology*, 96(3), 1592–1601.

Tanaka, J. W. & Taylor, M. (1991). Object categories and expertise: Is the basic level in the eye of the beholder? *Cognitive psychology*, 23(3), 457–482.

Tanaka, M. (2007). Recognition of pictorial representations by chimpanzees (pan troglodytes). *Animal cognition*, 10(2), 169–179.

Taylor, J., Hieber, L. L., & Ivry, R. (2012). Feedback-dependent generalization. *Journal of Neurophysiology*, 109(1), 202–215.

Theeuwes, J. (1992). Perceptual selectivity for color and form. *Attention, Perception, and Psychophysics*, 51(6), 599–606.

Tipper, S. P. (1985). The negative priming effect: Inhibitory priming by ignored objects. *Quarterly Journal of Experimental Psychology*, 37, 571–590.

Tipper, S. P. & Cranston, M. (1985). Selective attention and priming: Inhibitory and facilitatory effects of ignored primes. *The Quarterly Journal of Experimental Psychology Section A*, 37(4), 591–611.

Treisman, A. & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive psychology*, 12(1), 97–136.

Treisman, A. M. (2006). How the deployment of attention determines what we see. *Visual Cognition*, 14(4-8), 411–443.

Tseng, Y. w., Diedrichsen, J., Krakauer, J. W., Shadmehr, R., & Bastian, A. J. (2007). Sensory Prediction Errors Drive Cerebellum-Dependent Adaptation of Reaching. *Journal of Neurophysiology*, 98(1), 54–62.

Tukey, J. W. (1977). Exploratory data analysis. *Reading, MA: Addison-Wesley. Uncapher, M. R., Hutchinson, J. B., and Wagner, A. D. (*, 31(35), 12613–12628.

Tulving, E. & Schacter, D. L. (1990). Priming and human memory systems. *Science*, 247(4940), 301–306.

Tversky, B. (2010). Visualizing Thought. *Topics in Cognitive Science*, 3(3), 499–535.

Tversky, B., Suwa, M., Agrawala, M., Heiser, J., Stolte, C., Hanrahan, P., Phan, D., Klingner, J., Daniel, M.-P., & Lee, P. (2003). Sketches for design and design of sketches. *Human Behavior in Design: Individuals, Teams, Tools*, (pp. 79–86).

Uncapher, M., Hutchinson, J., & Wagner, A. (2011). Dissociable Effects of Top-Down and Bottom-Up Attention during Episodic Encoding. *Journal of Neuroscience*, 31(35), 12613–12628.

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

Von der Heydt, R., Peterhans, E., & Baumgartner, G. (1984). Illusory contours and cortical neuron responses. *Science*, 224(4654), 1260–1262.

von Helmholtz, H. (1925). Helmholtz's Treatise on Physiological Optics.

Walther, D. B., Chai, B., Caddigan, E., Beck, D. M., & Fei-Fei, L. (2011). Simple line drawings suffice for functional mri decoding of natural scene categories. *Proceedings of the National Academy of Sciences*, 108(23), 9661–9666.

Warrington, E. K. & Weiskrantz, L. (1974). The effect of prior learning on subsequent retention in amnesic patients. *Neuropsychologia*, 12(4), 419–428.

Watanabe, T., Náñez, J. E., & Sasaki, Y. (2001). Perceptual learning without perception. *Nature*, 413(6858), 844–848.

Watson, A. B. & Pelli, D. G. (1983). Quest: A Bayesian adaptive psychometric method. *Perception & Psychophysics*, 33, 2.

Wheeler, M. E. & Treisman, A. M. (2002). Binding in short-term visual memory. *Journal of Experimental Psychology: General*, 131(1), 48–64.

Wickham, H., Cook, D., & Hofmann, H. (2010). Graphical inference for infovis. *... and Computer Graphics ...*.

Wilken, P. & Ma, W. J. (2004). A detection theory account of change detection. *Journal of Vision*, 4(12), 1120–1135.

Williams, J. J. & Lombrozo, T. (2013). Explanation and prior knowledge interact to guide learning. *Cognitive psychology*, 66(1), 55–84.

Williams, M., Hong, S. W., Kang, M., Carlisle, N. B., & Woodman, G. F. (2013). The benefit of forgetting. *Psychonomic Bullein & Review*, 20(2), 348–355.

Williams, M., Hong, S. W., Kang, M.-s., Carlisle, N. B., & Woodman, G. F. (2012). The benefit of forgetting. *Psychonomic Bulletin & Review*, 20, 2.

Williams, M. & Woodman, G. F. (2012). Directed forgetting and directed remembering in visual working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 38(5), 1206–1220.

Witherspoon, D. & Allan, L. G. (1985). The effect of a prior presentation on temporal judgments in a perceptual identification task. *Memory & Cognition*, 13(2), 101–111.

Wu, S. & Keysar, B. (2007). The effect of information overlap on communication effectiveness. *Cognitive science*, 31(1), 169–181.

Yamins, D. L., Hong, H., Cadieu, C. F., Solomon, E. A., Seibert, D., & DiCarlo, J. J. (2014). Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proceedings of the National Academy of Sciences*, (pp. 201403112).

Yonas, A. & Arterberry, M. E. (1994). Infants perceive spatial structure specified by line junctions. *Perception*, 23, 1427–1427.

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