ATTENDING AND REMEMBERING THE EXTERNAL WORLD

A Thesis

Submitted to the Faculty

in partial fulfillment of the requirements for the

degree of

Doctor of Philosophy

in

Cognitive Neuroscience

by

Kirsten Ziman

DARTMOUTH COLLEGE

Hanover, New Hampshire

February 8, 2022

Examining Committee:

Jeremy R. Manning (Chair)

Theresa Desrochers

Emily Fynn

Tor Wager

F. Jon Kull
Dean of Graduate Studies

Abstract

The way we deploy our attention is continuously influencing our experience of the world around us and the aspects of our experience that we store into memory. In this dissertation, I consider visual attention and cognition in two respects: how visual attention interacts with memory, and how we can use our knowledge about visual cognition to inform the ways we visualize, interpret, and communicate scientific findings. First, I explore interactions between the high-level cognitive processes of attention and memory, investigating the effects of top-down feature-based and location-based attention on recognition memory. I find dissociable, but additive effects of feature-based and location-based attention on memory, operating on different timescales. As a complement to this behavioral finding, I also investigate what additional information about attention-memory interactions can be gleaned from pupillometry data. I find pupil dilations during initial stimulus viewing contain meaningful information about attention and memory processes: pupil dilation values during initial viewing of visual stimuli are predictive of later memory for attended stimuli only. Alternatively, pupil dilation values during recognition memory judgements are indicative of a subjective sense of memory, rather than veridical recognition memory processes (that is, they reflect what a person thinks they remember). After exploring visual attention and memory via behavioral and physiological measures, I shift my focus to ways we can utilize our knowledge of human vision and cognition to facilitate the intuitive interpretation of scientific data. I start by testing ways to reduce the human processing required for large corpuses of verbal data. Specifically, I analyze the capabilities of modern speech-to-text algorithms to automatically transcribe verbal data from psychology experiments, finding that they do so accurately, in ways that preserve pivotal aspects of the data relevant to studies of human memory. Next, I work to reduce the cognitive load of conceptualizing and interpreting high dimensional data. I combine powerful data reduction and alignment algorithms with visualization techniques to develop a new framework for intuitively visualizing and investigating high-dimensional data. In

the future, the combined findings and approaches in this dissertation may be used for a wide variety of applications, including pioneering efforts in computational psychiatry.

Acknowledgements

I would like to express my deepest gratitude to the many people who have supported me over the past five years. First, I would like to thank my advisor, Jeremy Manning, for taking a chance on me and for tirelessly encouraging my growth and professional development over the years. I also want to thank Peter Tse, Patrick Cavanagh, Theresa Desrochers, and all of the professors in the EPSCoR Attention consortium for their collective wisdom and support.

I also would like to thank all of the current and former members of the Contextual Dynamics Laboratory at Dartmouth, including Paxton Fitzpatrick, Lucy Owen, Andrew Heusser, Xinimng Xu, for their consistent feedback and friendship over the years. It has been a joy and an honor to work beside such dedicated, ambitious researchers. I also want to thank Sharif Saleki, for the loving support and countless late night conversations that have sustained me in these final years of graduate study, and for the pleasure of coauthoring research together.

Finally, I wan to express immense thanks to my friends and family, and especially to my parents for supporting my academic endeavors, and for helping me to obtain my first research assistant position over a decade ago. None of us ever imagined that this would be the end result.

Contents

List of Figures

List of Tables

Introduction

The goal of this dissertation is to study interactions between the high-level cognitive processes (attention and memory), and to apply what we know about human cognition to facilitate intuitive insights into scientific data.

Cognitive interactions: attention and memory

Attention and memory are two critical cognitive processes that, together, influence what we experience in each moment and what we remember later. Studying interactions between attention and memory can give us insights into learning, witness testimony, communication, and many other topics.

Importantly, we do not attend to or treat all types of remembered or incoming information equally; our flexible adaptation of our thinking and behaviors can change markedly with the specific concepts or tasks relevant to a given situation (Chun Turk-Browne 2007; Aly Turk-Browne 2017; Hardt Nadel 2009; Ranganath Ritchey 2012). Additionally, people vary individually with respect to the aspects of experience they notice, discriminate between, and act upon (sensory, social, emotional, etc.; (E. Hunt 1989), implying that the same physical (objective) experiences can give rise to very different perceived (subjective) experiences across people (Freeman Simoncelli 2011; Chang . 2018).

In some cases, we volitionally control our allocation of attention to various aspects of our experience, and in other cases our attention is deployed unconsciously and automatically (Jacoby 1992). Both types of attention may be expressed overtly, for example through eye movements (Hoffman Subramaniam 1995) or covertly, without any overt change in gaze position (Engbert Kliegl 2003). The neural bases of overt versus covert attention (Posner 1987; AR. Hunt Kingstone 2003) and volitional versus unconscious attention and their differential effects on memory (Dijksterhuis Aarts 2010) has

been explored in prior research.

This work has led to to the general consensus that sustained volitional attention enhances memory relative to unconscious attentional processes (Uncapher 2011; Turk-Browne 2013). However, many types of volitional attention exist, including attention to particular spatial locations and attention to particular visual features. How different types of volitional attention combine (or compete) to enhance memory has been an open question. I investigate this question in Aim 1 of this dissertation, using two complimentary behavioral experiments to explore the effects of spatial and feature-based attention on recognition memory.

Leveraging cognitive tendencies for data interpretation

Despite decades of research on visual attention, learning, and memory, the way we view, interpret, and communicate data does not always take advantage of what we know about human visual and cognitive systems. In recent years, researchers have begun working to incorporate principles from cognitive science into the *practice* of science and data interpretation itself. For example, research has suggested that object-based attention may contribute to false estimations of underlying data distributions while viewing standard bar graphs, but swarmplots (where all individual datapoints are displayed) clearly depict full data distributions while still yielding highly accurate estimations of the mean and highly accurate memory of estimated distribution means later on (Uddenberg 2016). So, in this case, by understanding how object based attention can bias perception, and by choosing instead to capitalize on our visual system's remarkably accurate perceptual averaging, we can make the scientifically informed decision to use swarmplots instead of barplots, when appropriate, to more accurately communicate our data. Tools that allow for easy creation of swarmplots and more transparent, visually intuitive data visualizations play an important role in facilitating this process (Waskom 2016).

Finding ways to intuitively visualize and interpret data will only become more important as time goes on. As we become inundated with “big data" from an increasingly technological world outside the lab, and complex data from increasingly naturalistic experiments within the lab, we will need to capitalize on our our brain’s natural strengths to help detangle sophisticated patterns and relationships in the data we collect.

Currently, two key obstacles are the sheer volume of data generated in the real world (and its in-lab approximations), and the inherent complexity (or high-dimensionality) of this data.

Scalability and data processing

Most psychology experiments require participants to respond to something they see, hear, or experience during the experiment. To make these responses computer-readable and easy to analyze, participants have traditionally responded in ways they rarely use in daily life. For example, in many classic memory studies (free recall studies), participants could not simply tell the experimenter what they remembered seeing, but might instead check a box or enter a typed responses on a keyboard. Now, as we move towards designing experiments that more closely mimic real life, participants are being encourage to remember what they’ve seen “out loud” (a more natural way of reporting memories).

While the spoken data format is more natural for participants to provide, and inherently more natural for a human listener to interpret, researchers cannot listen to, and individually interpret, the responses made by every participant in an experiment if the experiment is to be conducted on a large scale. The amount of data is simply too vast. To do so would require immense time and resources processing verbal data – first taking audio recordings of each participant’s responses for the duration of the experiment, then meticulously transcribing those audio recordings by hand after the experiment ended, while

attempting to standardize transcription practices across transcribers and studies. Now, however, with the advent of speech-to-text-engines, we have the technology to process natural verbal responses quickly, easily, and automatically. In the first portion of Aim 2, I test the efficacy of automatic speech-to-text transcription methods for psychological purposes and present tools to automatically convert vast amounts of verbal response data into computer readable formats, tailored for memory experiments.

Complexity and data analysis

In addition to scalability, there is also the issue of data complexity. Conceptualizing high-dimensional data is difficult, since our visual system is optimized for three-dimensional objects. Instead of fighting our natural inclinations to try and imagine higher-dimensional patterns, we can work to display information in lower dimensional spaces in ways that capitalize on our visual system's natural strengths and dimensional capacities (including processing information along five dimensions: three spatial dimensions, one color dimension, and one temporal dimension). In this dissertation, I use data visualization and manipulation techniques to render complex, high-dimensional data more intuitively interpretable.

Modern data visualizations date back to at least the 16th century, when early data pioneers began to develop the sorts of accurate maps and diagrams we might still recognize today (Friendly 2006; Tufte Graves-Morris 1983). Visualizations can reveal deep insights and intuitions about geometric structure and patterns in complex datasets by capitalizing on the human visual system's ability to quickly and efficiently extract meaning and structure from highly complex visual information (Uddenberg 2016). This may be especially true of high-dimensional datasets, where different dimensions or features of the data can interact in complex ways that may not be immediately obvious through standard summary statistics.

The classic example, Anscombe's Quartet (Anscombe 1973), provides an excellent illustration of the potential for summary statistics to mislead in the absence of visualization. Anscombe's Quartet comprises four datasets that share a common statistical profile. The datasets seem highly similar, because they are exactly equal along several common summary measures (mean, variance, trend lines). However, plotting the datasets and comparing them visually reveals that they differ substantially in structure. Whereas low-dimensional datasets like those in Anscombe's Quartet can be easily plotted, it is not always obvious how to visualize high-dimensional datasets (e.g. with greater than 3 dimensions) in a similarly intuitive way.

One important class of techniques, collectively referred to as *dimensionality reduction algorithms* have been developed over the past half-century to map high-dimensional data onto lower-dimensional representations that are easier to manipulate and visualized. Examples of these algorithms include Principal Components Analysis (PCA) (Pear-son 1901), Probabilistic Principal Components Analysis (PPCA) (Tipping Bishop 1999), Independent Components Analysis (ICA) (Jutten Herault 1991; Comon 1991), Multidi-mensional Scaling (MDS) (Torgerson 1958), and *t*-Distributed Stochastic Neighbor Em-bedding (t-SNE) (van der Maaten Hinton 2008). Each of these algorithms provides a slightly different means of obtaining a low-dimensional representation of the original high-dimensional dataset in a way that preserves as many of the geometric properties of the original high-dimensional data (e.g. the overall covariance structure of the data, data grouping, etc.) as possible, within a low-dimensional space. As such, these dimensionality reduction algorithms are especially helpful for visualizing high-dimensional data.

A second class of helpful algorithms, which draw inspiration from the Procrustean transformation (Schönemann 1966), provide techniques for manipulating and aligning different high-dimensional datasets. These algorithms compute the affine transformations (i.e. translation, reflection, rotation, and scaling) that bring one

trajectory into alignment with another (in terms of minimizing the mean squared Euclidean distances between the corresponding points). The hyperalignment algorithm (Haxby 2011) and the Shared Response Model (SRM) (Chen 2015) extend this technique to find a common set of transformations that bring many (more than two) high-dimensional trajectories into common alignment. These alignment algorithms are especially powerful in that they can be used to manipulate and compare high-dimensional datasets that do not share the same original dimensions (or features). For example, they can be used to compare brain patterns from different people, observations collected via different modalities, and so on.

In the second portion of Aim 2, I combine dimensionality reduction algorithms, data alignment algorithms, and visualization techniques to create a new framework that capitalizes on human perceptual skills to facilitate intuitive insights into high-dimensional data. Additionally, I implement a comprehensive, easy-to-use toolbox for exploring data this way.

Dissertation Aims

Aim 1: Cognitive interactions: Attention & memory

My first aim (Chapters 1 and 2) explores how volitional attention affects our memory for the things we experience. Specifically, I investigate how volitional covert attention (i.e. attention paid "out of the corner of your eye") to specific visual stimuli influences later memory for those stimuli *and* for other nearby stimuli.

For this purpose, I designed a novel task which combines a classic attention experiment paradigm (covert attention) with a classic memory experiment paradigm (recognition memory). Further, I utilized composite stimuli (overlaid pairs of images) within the task, allowing me to separately quantify the effects of spatial attention versus feature-based attention on later recognition memory. This experiment design allowed me to explore how recognition memory changes for images that are presented under identical conditions (externally, and under the same gaze conditions of the participant) but vary in terms of the amount of attention they receive during initial viewing (spatial and feature-based attention). After analyzing behavioral data from this experiment, I further explored physiological correlates of the interactions between attention and memory by analyzing pupil dilation data collected during the experiment.

Aim 2: Leveraging cognitive tendencies for data interpretation

My second aim (Chapters 3 and 4) focuses on leveraging what we know about the human visual system to facilitate more intuitive interpretation and communication of scientific data. This aim is particularly relevant as we increasingly work with naturalistic

datasets. Two aspects that make this data unintuitive to work with are that data collected in more naturalistic contexts is often complex and difficult to process (making it difficult to scale up experiments to larger numbers of participants) and that this data, once processed, is often high-dimensional, making it tedious and difficult to analyze. As such, my second aim implements computational approaches for easily processing and interpreting complex and high-dimensional datasets.

First, I showcase the feasibility and research impacts of automating the otherwise time-intensive, tedious, and subjective task of transcribing verbal responses recorded during psychology experiments. I analyzed the ability of a state-of-the-art speech recognition algorithm to transcribe verbal response data recorded during a verbal free recall task. In addition to analyzing the algorithm's over all accuracy, I probed its ability to transcribe responses in a way that preserved important, psychologically relevant statistical properties of the dataset.

Second, I integrated powerful data reduction, alignment, and visualization approaches in order intuitively manipulate and visualize high-dimensional data. I use these combined approaches to explore the underlying structure in multiple high-dimensional datasets (including brain and behavioral data). I find that these integrated methods (which I organized into an open-source Python software package, HyperTools) make it easy reveal and capitalize on critical structure and relationships in high-dimensional data, and can be applied to a wide variety of data types (timeseries, categorical, textual) collected across modalities (behavioral, physiological, neural).

Chapter 1

Spatial and category-level attention affect memory at different timescales

Ziman K., Lee M. R., Martinez A. R., Adner, E. D. & Manning J. R. (2020). Feature-based and location-based volitional covert attention are mediated by different mechanisms and affect memory at different timescales. *PsyArXiv: 10.31234/osf.io/2ps6e.*

1.1 Introduction

Our brain's cognitive systems detect and exploit patterns in our prior and ongoing experiences, enabling us to function and adapt in an ever-changing world. However we do not attend to or treat all types of remembered or incoming information equally, and our ability to flexibly adapt our thinking and behaviors can vary markedly with the specific set of concepts or tasks relevant to a given setting or situation (Chun Turk-Browne 2007; Aly Turk-Browne 2017; Hardt Nadel 2009; Ranganath Ritchey 2012). There is also substantial variability across people with respect to which aspects of experience (sensory, social, emotional, etc.) are noticed, discriminated between, and acted upon (E. Hunt 1989). This implies that the same physical (objective) experience may give rise to very differ-

ent perceived (subjective) experiences across people (Freeman Simoncelli 2011; Chang 2018).

The aspects of our experience we attend may be under our volitional control or may be unconscious or automatic (Jacoby 1992). Both volitional and unconscious attention may be expressed overtly, for example through intentional eye movements (Hoffman Subramaniam 1995) or covertly, without any volitional physical change (Engbert Kliegl 2003). Prior work has explored the similarities and differences in the neural basis of overt versus covert attention (Posner 1987; AR. Hunt Kingstone 2003) as well as the behavioral and neural underpinnings of volitional versus unconscious attention (Dijksterhuis Aarts 2010) and their differential effects on memory. There is a general consensus that sustained volitional attention enhances memory relative to unconscious attentional processes (Uncapher 2011; Turk-Browne 2013). However, volitional attention takes many forms, such as attention to particular spatial locations or attention to particular visual features. How different *types* of volitional attention combine (or compete) to enhance memory remains an open question. Volitional covert attention is of particular interest in that it allows us to dynamically and intentionally manipulate our experience, even when our sensory input remains largely static (i.e., constant physical stimuli, retinal image, etc. Yi 2006; O'Craven 1999).

Here we examine the ways different types of volitional covert attention interact to affect memory. We designed an experimental paradigm (following Posner 1980) that asked participants to attend to a series of presented composite image pairs while keeping their gaze fixed on a central point. The image pairs comprised a left and right image, each constructed by blending an image of a face and scene. The stimuli and presentation durations were constant across the two experiments, but the experiments differed in how often we asked participants to change the focus of their attention with respect to image category (face versus scene) and image location (left versus right). After the participants

attended to a series of images, we used a recognition memory test to assess which aspects of the presented images had been encoded into memory. In both experiments we found that the images participants covertly attended to were better recognized than other images, supporting the notion that attention enhances memory encoding (i.e., they rated attended images as more familiar than unattended images Yonelinas 2002). After maintaining the focus of attention to a single image category and location (Sustained Attention Experiment), participants also recognized the attended-category image at the unattended location, and (to a lesser extent) the unattended-category image at the attended location. After more rapidly varying their focus of attention (Variable Attention Experiment), participants showed a similar boost in recognition for the unattended-category image at the attended location, but they did not recognize images at the unattended location. This suggests that participants were able to shift the location of their covert attentional focus more rapidly than they were able to shift their focus of covert attention to stimulus features. We also found differences in the timecourses of these memory effects, suggesting that the impact of location-based attention on memory persists on the order of several seconds longer than the impact of feature-based attention.

1.2 Materials and methods

We ran a total of 113 participants in two covert volitional attention experiments (Fig. 1.1). The experiments were similar, except in how often we cued participants to change the focus of their attention. All code and documentation pertaining to our experiments and analyses, along with the experimental stimuli and data, may be downloaded from http://www.github.com/ContextLab/attention-memory-task.

Initially, we ran a total of 60 participants (30 in each experiment), but following advice from reviewers, we aimed to double our sample. When preparing to collect our second sample, we noticed minor bugs in our experiment code, which we corrected (for details,

see documentation). We then collected data from 30 additional participants for the Sustained Attention Experiment and proceeded to collect replication data for the Variable Attention Experiment. After collecting 23 participants for the Variable Attention Experiment, the COVID-19 pandemic halted in-person laboratory testing. As such, the replication cohort for the Variable Attention Experiment has 23 participants (rather than our target sample size of 30). To confirm that the minor adjustments made to our experiment code did not influence participants' performance, we conducted analyses separately on the data collected before and after implementing those adjustments (see *Supporting Materials*). When conducted separately on each sample, the analyses reported in the text of this manuscript yield compatible results across both the initial and the replication groups of each experiment.

Participants

Sustained Attention Experiment

Sixty students, enrolled in one or more undergraduate psychology courses at Dartmouth College, participated in the Sustained Attention Experiment for course credit. The participants' ages ranged from 18–21; 42 participants were female, 17 were male, and one identified as gender non-binary. All participants had self-reported normal or corrected-to-normal vision, memory, and attention. Participants gave written consent to enroll in the study under a protocol approved by the Committee for the Protection of Human Subjects at Dartmouth College.

Variable Attention Experiment

Fifty-three individuals participated in the Variable Attention Experiment for either cash (n = 13) or course credit (n = 40; students enrolled in one or more undergraduate psychology courses at Dartmouth College). The participants' ages ranged from 18–23; 34 participants were female and 19 were male. All participants had self-reported normal or

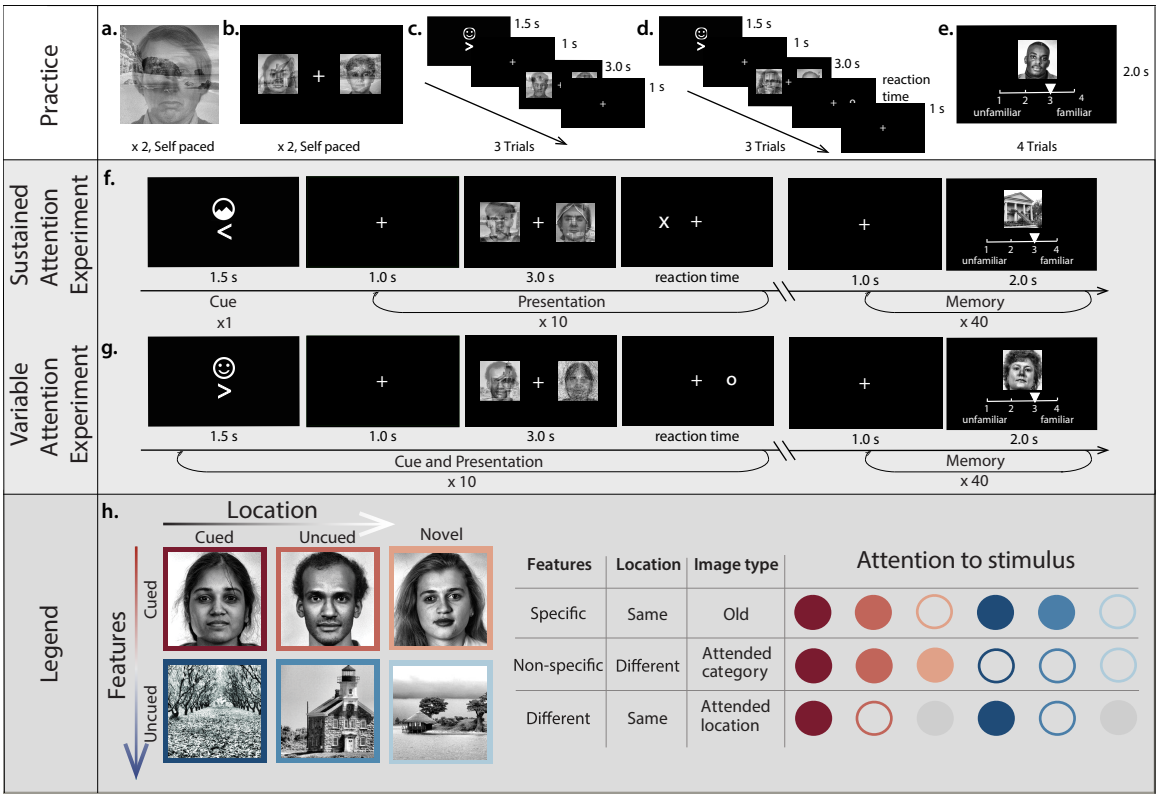


Figure 1.1: Covert attention & recognition memory paradigm.

a. Composite face/scene image. **b.** A single pair of composite images and a central fixation cross. **c.** Timecourse of an attention cue practice trial (3 total). **d.** Timecourse of a practice trial for the reaction time task (3 total). **e.** Familiarity judgement practice trial (4 total). **f.** Timecourse of one Sustained Attention Experiment presentation phase and memory phase block. **g.** Timecourse of one Variable Attention Experiment presentation phase and memory phase block. **h.** Breakdown of presented (old) and novel stimuli that matched (or did not match) the focus of participants' feature-based attention and/or location-based attention. Filled colored circles on the right denote that the given stimulus (matching the corresponding color; images shown on the left) satisfied the indicated property. Open circles denote that the given stimulus did *not* satisfy the indicated property. Gray circles denote properties that are undefined for the given stimulus.

corrected-to-normal vision, memory, and attention. Participants gave written consent to enroll in the study under a protocol approved by the Committee for the Protection of Human Subjects at Dartmouth College.

Stimulus selection and presentation

Participants viewed photographs of faces, scenes, and composite images each comprising an equal blend of one face image and one scene image. The pool of 360 face images included photographs of adult human male and female faces selected from the FERET database (Phillips 1998). The pool of 360 scene images included photographs of indoor and outdoor scenes selected from the SUN database (Xiao 2010). The images we used from both databases came from a stimulus subset that was manually curated by Megan deBettencourt (personal communication). All images were resized to 256 × 256 pixels, converted to greyscale, and processed so that every image was matched for mean contrast and intensity. We selected 20 face images and 20 scene images from the stimulus pool to use in the instructional and practice phases of the experiments (Fig. 1.1a–e).

In addition to the face and scene images, we presented (in white) attention cues to direct the participant's focus of attention. The attention cues comprised a stylized icon of a face or mountain peaks, directing attention to the face or scene component of the images, respectively; and a left- or right-facing angled bracket, directing attention to the left or right image, respectively (e.g., Figs. 1.1c, d, f, and g).

Both experiments were conducted in a sound- and light-attenuated testing room containing a chair, desk, and 27-inch iMac desktop computer (display resolution: 2048 × 1152). The participant sat in the chair and rested their chin on a chin rest located 60 cm from the display. The active portion of the display screen occupied 52.96° (width) and 31.28° (height) of the participant's field of view from the chin rest. Stimuli were sized to occupy 6.7° (width and height) of the participant's field of view from the chin rest. We maintained

a black background (with any text displayed in white) throughout the experiment.

Eyetracking

We recorded participants' eye gaze positions using a desk-mounted video-based eyetracker with a spatial resolution of 0.1° visual angle root mean squared error and a sampling rate of 30 Hz (Eye Tribe, The Eye Tribe, Copenhagen, Denmark). We calibrated the eyetracker using a 9-point gaze pattern. As described below, we re-calibrated the eyetracker at regular intervals throughout the experiments to protect against camera drift.

Experimental paradigm

Both experimental paradigms comprised a practice phase followed by a series of eight task blocks. Each task block was in turn comprised of a presentation phase and a memory phase. The practice and presentation phases differed across the two experiments, and the memory phases were identical across the two experiments. Both experiments were implemented using PsychoPy (Peirce 2019).

Practice phase

Several participants in pilot versions of our experiments reported that they found it difficult to modulate the focus of their attention quickly on command. We therefore designed a practice sequence to orient the participant to the process of quickly modulating the focus of their attention without moving their eyes. The experimenter remained in the testing room throughout the practice phase and answered any questions about the experiment. The practice sequence builds up incrementally to provide a gradual on-ramping for the participant prior to beginning the main experimental tasks that we focused on in our analyses.

Practice shifting the focus of feature-based attention to elements of a single composite image. At the start of the practice phase, we instructed the participant to look at a single composite (face-scene blend) image at the center of the screen, and to try to bring the face component of the image into greater focus by attending to it (Fig. 1.1a). After pressing a button on the keyboard to indicate that they had done so, we displayed a second composite image and instructed the participant to bring the scene component of the new composite image into focus. Again, they pressed a button to indicate that they had done so.

Practice shifting the focus of feature-based and location-based attention while viewing two composite images. Next, we asked the participant to stare at a fixation cross presented in the center of the screen while two composite images were displayed on the left and right side of the screen, respectively. We first instructed the participant to attend to the scene component of the left image without moving their eyes. Participants practiced shifting their attention, and they pressed a button on the keyboard to indicate that they had done so. We then displayed a second pair of composite images and instructed the participant to attend to the face component of the right image. Again, the participant shifted their attention in a self-paced manner, and pressed a button to indicate when they had successfully done so (Fig. 1.1b).

Practice sustaining feature-based and location-based attention over a series of composite image pairs. We asked participants in the Sustained Attention Experiment to practice holding their focus of feature-based and location-based attention constant (to the face component of the right image) while viewing a series of three composite image pairs presented in succession (Fig. 1.1c).

Practice varying feature-based and location-based attention over a series of composite image pairs. We asked participants in the Variable Attention Experiment to

practice varying their focus of feature-based and location-based attention while viewing a series of three composite image pairs presented in succession (Fig. 1.1c).

Practice reaction time probe. After practicing modulating their focus of attention to a series of composite image pairs, we introduced a reaction time probe after each image presentation, whereby we presented either an × or ∘ on either the left or the right of the screen (Fig. 1.1d). We asked the participant to press the 1 key as quickly as possible when they saw an ×, or the 3 key as quickly as possible when they saw an ∘. We did not impose a time limit on their responses, other than asking participants to respond as quickly as they were able. Participants practiced three trials of modulating their focus of attention to a pair of composite images (3 s), and reacting as quickly as possible to the × or ∘ symbol presented after each composite image pair. The reaction time probe was intended to keep participants continually engaged in modulating the focus of their attention.

Practice recognition memory task. Finally, we asked the participant to practice reporting familiarity on a recognition memory task (Fig. 1.1e). We presented a single face or scene image at the center of the screen, and asked them to press a button to indicate how "familiar" the image seemed: 1 (very confident they had not seen the image), 2 (somewhat confident they had not seen the image), 3 (somewhat confident they had seen the image), or 4 (very confident that they had seen the image). We instructed the participant to go with their "gut reaction" in the event that they were unsure of how to respond. We allowed the participant up to 2 s to provide their response. We gave participants a total of four practice images to rate.

After completing the practice phase of the experiment, the participant read the instructions for the task blocks (described next). The experimenter gave participants a chance to ask any remaining questions about the experiment. After answering the participant's questions, the experimenter calibrated the eyetracker and exited the testing room.

Task blocks

During each task block we asked the participant to modulate their attention while viewing a series of 10 composite image pairs (each followed by a reaction time probe), and then we tested the participant's memory using 40 familiarity judgements. Each participant completed a total of eight task blocks.

Sustained Attention Experiment: presentation phase. Participants viewed an attention cue (1.5 s) instructing them to attend to either the face or scene component of either the left or right images in each to-be-viewed composite pair. Next we displayed 10 composite images in succession (each preceded by a fixation cross and proceeded by a reaction time probe). Complete timing information is displayed in Figure 1.1f. All possible attention cue pairs appeared exactly twice across the eight task blocks.

Variable Attention Experiment: presentation phase. Participants viewed a succession of 10 attention cues (1.5 s), each followed by a fixation cross (1 s), composite image (3 s), and a reaction time probe (Fig. 1.1g). The attention cues were selected pseudorandomly across trials within each block, with the constraints that no single attention cue pair could appear more than three times across the 10 composite image pairs within a single task block.

Memory phase. After the presentation phase of each task block, we asked the participant to rate the familiarity (on a 1–4 scale, as during the practice phase) of a succession of face and scene images. Each image was preceded by a 1 s fixation cross, and participants had up to 2 s to input their rating of each image. Participants made a total of 40 familiarity judgements, about 20 face images and 20 scene images. Of these, 20 of the images (10 faces and 10 scenes) were drawn randomly from the (attended and unattended) composite images that the participant had viewed during the presentation phase. The remaining 20 images (10 faces and 10 scenes) were novel images that the participant had not encoun-

tered during any part of the experiment. At the end of each memory block, the participant was given the opportunity to take a short break. When they were ready to continue with the next task block, they indicated their readiness to the experimenter. The experimenter then entered the testing room, re-ran the eyetracker calibration sequence, and exited the testing room prior to the next task block.

1.3 Results

We ran two volitional covert attention experiments; in the Sustained Attention Experiment we asked participants to *sustain* the focus of their attention over a succession of 10 stimulus presentations per block whereas in the Variable Attention Experiment we asked participants to *vary* their focus of attention with each new stimulus (also for a total of 10 stimulus presentations per block). Each stimulus comprised a pair of composite images (one on the left and one on the right side of the display), where each composite comprised an equal blend of a unique face and a unique scene image. We followed the presentation phases of each experimental block with a memory phase, where participants performed a recognition memory task by rating the familiarity of previously experienced and novel face and scene images (see *Materials and methods*, Fig. 1.1).

We considered each image we probed in the memory phase with respect to whether (and how) the participant had experienced and/or attended to that image during the presentation phase (Fig. 1.2). For example, if a given face image had been presented as part of the composite image on the right of the screen following an `attend to the face on the right` instruction, then that face image would have been the focus of the participant's attention during one stimulus presentation. The face image on the left (unattended side) during that trial would have been captured by feature-based attention, but not by location-based attention; and the scene image (unattended category) on the right during that trial would have been captured by location-based attention, but not by feature-based

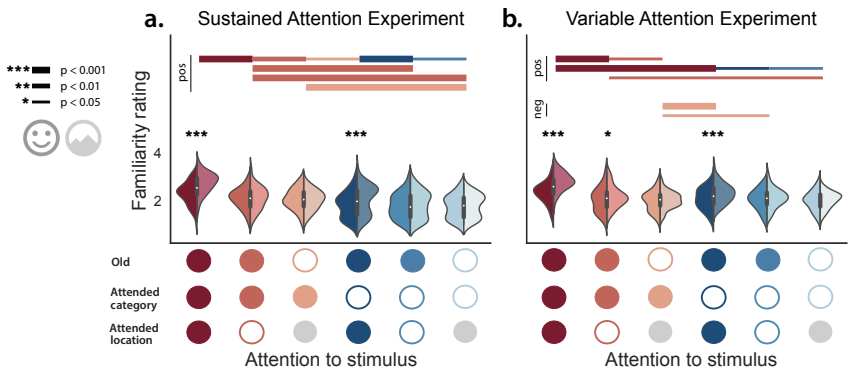


Figure 1.2: Familiarity ratings for attended, unattended novel stimuli.

Each split violin plot displays the distribution of within-participant average familiarity ratings given to faces (left, darker colors) and scenes (right, lighter colors) during the memory phases of each experiment. As shown in the legend (bottom), the colors indicate whether each image had been viewed during the presentation phase (`old`) or not; whether the given images matched the `attended category`; and/or whether the given images matched the `attended location`. The colored lines above each set of violin plots denote statistical differences (**pos**itive or **neg**ative differences in mean, collapsing over image category, assessed via two-tailed *t*-tests) between the distributions centered on the endpoints of each line. The line thicknesses denote *p*-values as indicated in the legend. Asterisks denote differences between the face versus scene distributions (assessed via two-tailed *t*-tests). Panel **a.** displays results from the Sustained Attention Experiment and panel **b.** displays results from the Variable Attention Experiment. Appendix A displays these results broken down by participant cohort.

attention. The scene image on the left during that trial remained outside of the focus of both feature-based and location-based attention. In this way, we categorized each of the images that participants experienced during the presentation phases of each experiment in terms of whether they fell under the scope of feature-based and/or location-based attention. We also categorized novel face and scene images that we asked participants to judge during the memory phase (i.e., images that the participant hadn't seen before) as belonging to the attention-cued or uncued category. These novel images were intended to serve as a baseline for comparison. For example, we sought to measure how people rated the familiarity of new images in general. We also used differences in familiarity ratings between novel images of the attention-cued versus uncued categories to evaluate potential *prospective* affects of modulating the focus of feature-based attention.

Participants in both experiments rated stimuli they attended as more familiar than unattended or novel stimuli (Fig. 1.2; Sustained Attention Experiment: $t(59) = 13.42, p = 1.39 \times 10^{-19}, d = 1.73$; Variable Attention Experiment: $t(52) = 12.40, p = 3.70 \times 10^{-17}, d = 1.70$). Participants in both experiments also rated attended scene stimuli as more familiar than attended face stimuli (Sustained Attention Experiment: $t(59) = 7.69, p = 1.85 \times 10^{-10}, d = 0.99$; Variable Attention Experiment: $t(52) = 7.08, p = 3.63 \times 10^{-9}, d = 0.97$). These findings indicate that aspects of experience that are captured by the focus of attention are easier to recognize later, and that attention may preferentially benefit memory for some aspects of experience to a greater extent than other aspects of experience.

Participants in the Sustained Attention Experiment experiment also rated unattended images that matched the attended category as more familiar than unattended images that did not match the attended category (*incidental feature-based attention*; $t(59) = 7.17, p = 1.40 \times 10^{-9}, d = 0.93$). This pattern did not hold for participants in the Variable Attention Experiment ($t(52) = 0.44, p = 0.66, d = 0.06$). Participants in both experiments displayed an incidental effect of location-based attention, whereby they rated unattended-category images at the attended location as more familiar than unattended-category images at the unattended location (Sustained Attention Experiment: $t(59) = 4.32, p = 6.05 \times 10^{-5}, d = 0.56$; Variable Attention Experiment: $t(52) = 2.57, p = 0.01, d = 0.35$). This indicates that aspects of our experience that incidentally overlap with the focus of our location-based attention are easier to recognize later, even if those aspects do not fall within the intentional focus of our feature-based attention.

We defined the *memory benefit of feature-based attention* as the difference in mean familiarity ratings of attended stimuli and unattended stimuli from the attended category versus the mean familiarity ratings of unattended stimuli from the unattended category. Similarly, we defined the *memory benefit of location-based attention* as the difference in mean familiarity ratings of attended stimuli and unattended stimuli from the attended location

versus the mean familiarity ratings of unattended stimuli from the unattended location. Participants in the Sustained Attention Experiment displayed a larger memory benefit of feature-based attention than location-based attention ($t(59) = 3.25, p = 1.91 \times 10^{-3}, d = 0.42$), whereas participants in the Variable Attention Experiment displayed a larger memory benefit of location-based attention ($t(52) = 1.70, p = 9.99 \times 10^{-2}, d = 0.23$). This appeared to be driven in part by a response bias towards attended-category images. For example, relative to the most recent attention cue, participants in the Sustained Attention Experiment rated attended-category novel images as more familiar than unattended-category novel images ($t(59) = 7.04, p = 2.35 \times 10^{-9}, d = 0.91$), whereas participants in the Variable Attention Experiment responded similarly to attended-category and unattended-category novel images ($t(52) = 0.23, p = 0.82, d = 0.03$). The size of this attended-category response bias for novel images was reliably smaller than the memory benefits of feature-based attention in both experiments (Sustained Attention Experiment: $t(59) = 7.59, p = 2.68 \times 10^{-10}, d = 0.98$; Variable Attention Experiment: $t(52) = 6.74, p = 1.31 \times 10^{-8}, d = 0.93$). This indicates that tuning the focus of feature-based attention with respect to how ongoing experience is encoded in memory takes longer than the duration of a single stimulus presentation (including fixation, roughly 5.5 s). Further, once feature-based attention comes "online," it affects how *new* stimuli are processed.

We also compared the relative sizes of the memory benefits of feature-based attention and location-based attention across the two experiments. Participants in the Sustained Attention Experiment displayed a larger difference between these two effects than did participants in the Variable Attention Experiment ($t(111) = 3.48, p = 7.24 \times 10^{-4}, d = 0.66$). Further, the difference in mean familiarity ratings of attended-location images versus unattended images was greater for participants in the Variable Attention Experiment ($t(111) = 2.32, p = 0.02, d = 0.44$). The memory benefit of location-based attention reliably distinguished between attended-category versus unattended category images for participants in both experiments (Sustained Attention Experiment: $t(59) = 9.50, p = 1.68 \times 10^{-13}, d =$

1.23; Variable Attention Experiment: $t(52) = 9.72, p = 2.76 \times 10^{-13}, d = 1.33$). This indicates that tuning the focus of location-based attention with respect to how ongoing experience is encoded into memory occurs faster than the duration of a single stimulus presentation.

We also compared reaction times to probes from the attended versus unattended side. Participants in both experiments responded slightly faster on average to attended-side probes, although the differences did not reach significance (Sustained Attention Experiment: $t(59) = 1.03, p = 0.31, d = 0.13$; Variable Attention Experiment: $t(52) = 0.65, p = 0.52, d = 0.09$). Participants in the Sustained Attention Experiment displayed a slightly larger (and not statistically significant) difference in attended-side versus unattended-side reaction times than participants in the Variable Attention Experiment ($t(111) = .87, p = 0.38, d = 0.20$). Because none of these reaction time effects were statistically reliable, we are unable to draw meaningful conclusions about reaction times to the attention probes on the attended versus unattended side, nor are we able to draw meaningful conclusions about differences in those reaction times that might be due to sustained versus variable attention.

Given that participants in both experiments rated attended-location images from both the attended and unattended category as more familiar than unattended-location images, we wondered whether this was truly a consequence of covert attention, or whether it might be driven solely by where participants were looking. In other words, we sought to distinguish the memory effects of the focus of covert attention from the focus of visual gaze. We used eyetracking to measure the horizontal positions of the participants' visual fixations while each pair of composite stimuli appeared onscreen (Fig. 1.3). For 94% of presentation trials (Sustained Attention Task: $\frac{4531}{4800}$ trials; Variable Attention Task: $\frac{3972}{4240}$ trials), we collected viable gaze data for analysis. In the remaining trials, participant movement or other technical issues corrupted the gaze data. We found that, though they generally followed

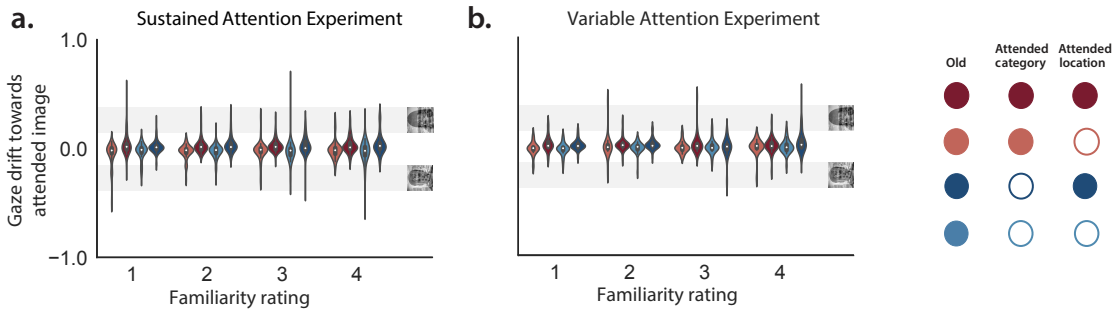


Figure 1.3: Horizontal coordinates of visual fixation during testing.

Each violin plot displays a distribution of participant-wise average horizontal gaze positions relative to a presented image. The coordinates have been normalized such that a value of 1.0 denotes the furthest onscreen coordinate from the central fixation point *towards* the direction of the given image, and -1.0 denotes the furthest onscreen coordinate from the central fixation point *away from* the direction of the given image. The gray bars in each panel mark the boundaries of the presented images. The violin plots are broken down by participants' familiarity ratings during the memory phases of each experiment, and by each image's relation to the attention cue while that image appeared onscreen. Panel **a.** displays fixation data for the Sustained Attention Experiment and Panel **b.** displays fixation data for the Variable Attention Experiment. Appendix A displays these results broken down by participant cohort.

our instruction to keep their gaze fixed on the center of the screen, participants in both experiments exhibited a slight tendency to look towards the side of the screen they were attending to (Sustained Attention Task: $t(59) = 3.61, p = 6.31 \times 10^{-4}$; Variable Attention Task: $t(52) = 2.86, p = 6.06 \times 10^{-3}$). When we limited our analyses to trials with viable gaze data where participants successfully maintained fixation on the center of the screen within 4.28° visual angle (i.e., between the rightmost boundary of the left image and the leftmost boundary of the right image) we observed (numerical) memory benefits of feature-based attention in both experiments (Sustained Attention Experiment: $t(59) = 4.73, p = 1.96 \times 10^{-5}, d = 0.67$; Variable Attention Experiment: $t(52) = 0.64, p = 0.52, d = 0.09$) and location-based attention in the Sustained Attention Experiment, but not the Variable Attention Experiment (Sustained Attention Experiment: $t(59) = 0.67, p = 0.51, d = 0.09$; Variable Attention Experiment: $t(52) = -0.79, p = 0.43, d = 0.11$). These analyses indicate that although the focus of covert attention can affect where people are looking, covert

attention affects memory encoding beyond what may be accounted for by gaze position alone.

The preceding results indicate that volitional covert attention affects memory, and that feature-based and location-based covert attention may come online at different timescales. We next sought to examine how long the memory effects of feature-based and location-based covert attention persist. To gain additional insights into the timecourse of the impact of feature-based and location-based attention on memory, we used a sliding window analysis to measure how participants' familiarity ratings of attended, partially attended, unattended, and novel stimuli varied over the duration of the memory phases of both experiments. During each block within the memory phases of each experiment, participants made familiarity judgements about a total of 40 images (20 old images and 20 novel images). We computed average familiarity ratings for images appearing in each position of the memory stimulus sequence (positions 1–40), as a function of how (or whether) they were attended during the preceding presentation phase (old images) or whether they matched the category of the most recent attention cue (novel images), for overlapping 20-image sliding windows (Figs. 1.4, 1.5, and 1.6). We first calculated these values for each participant, and then we averaged across participants.

Throughout the memory phase, participants from both experiments rated attended images as more familiar than category-matched unattended and novel images (Fig. 1.4, top row). This indicates that the memory benefits of attending to specific stimulus features (versus category-general features) persist for at least as long as the duration of the memory phase (2 min). Early in the memory phase, participants in the Variable Attention Experiment also rated category-matched unattended images as more familiar than novel images. This pattern did not hold for participants in the Sustained Attention Experiment. The short-lived boost in familiarity for attended-category images at the unattended location following variable (but not sustained) attention suggests that the timecourse of

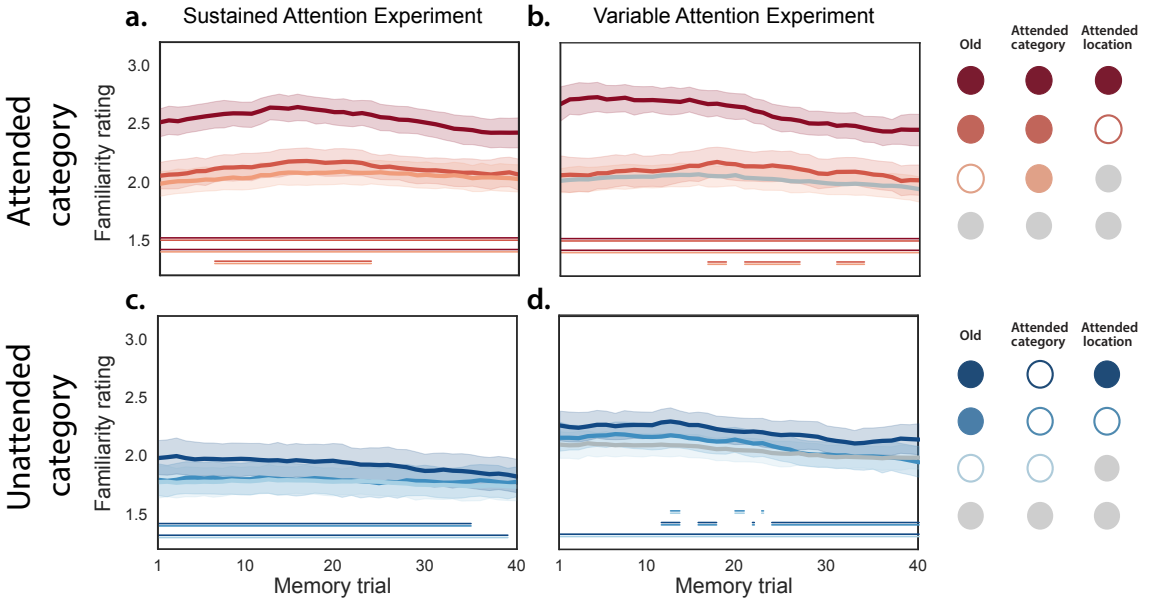


Figure 1.4: Familiarity ratings over time for images that matched the cued or uncued image category.

Each curve reflects the average familiarity ratings for attended, unattended, and novel images (denoted in the legends on the right) within a succession of overlapping 20-image sliding windows. Error ribbons denote 95% confidence intervals, computed across participants. Panels **a.** and **c.** display results from the Sustained Attention Experiment and Panels **b.** and **d.** display results from the Variable Attention Experiment. The paired horizontal lines at the bottom of each panel denote timepoints when the given pair of curves was statistically distinguishable (i.e., the topmost line color was statistically greater than the bottommost line color at $\alpha = 0.05$, via a paired two-tailed t-test.) The gray lines in Panels b and d reflect familiarity ratings of novel stimuli from both the most recently attended and unattended image categories. Appendix A displays these results broken down by participant cohort.

feature-based attention's impact on memory depends on how long the focus of feature-based attention has been held.

Participants in both experiments also rated unattended-category images at the attended location as more familiar than unattended-category images at the unattended location and novel unattended-category images (Fig. 1.4, bottom row). This pattern persisted throughout the memory phases of both experiments. This supports the notion that location-based attention enhances memory in a (partially) feature-independent way. Participants in the Variable Attention Experiment also rated unattended-category images at the unattended location as (numerically) more familiar than novel images (though this pattern was only statistically reliable towards the middle of the memory phase). This pattern did not hold for participants in the Sustained Attention Experiment. Participants in the Sustained Attention Experiment rated unattended-category images (regardless of location or novelty status) as less familiar than participants in the Variable Attention Experiment (e.g., compare heights of the lines in the bottom left versus bottom right panels of Fig. 1.4). This suggests that, after sustained feature-based attention to a particular image category, the encodings of stimuli from the unattended category are suppressed. This suppression effect appears to persist at least as long as the memory phase of the experiment (2 min). Further, the magnitude of this suppression effect appears to become stronger following longer-term sustained (versus shorter-term variable) feature-based attention.

We next examined how participants' familiarity ratings varied over the duration of the memory phases of each experiment as a function of the attended location. Participants in the Sustained Attention Experiment rated attended-category images as more familiar, at both the attended and unattended locations (Fig. 1.5, left panels). In contrast, participants in the Variable Attention Experiment rated attended-category images as more familiar than unattended-category images only at the attended locations (Fig. 1.5, right panels). The primary difference in familiarity ratings between Sustained Attention Experiment

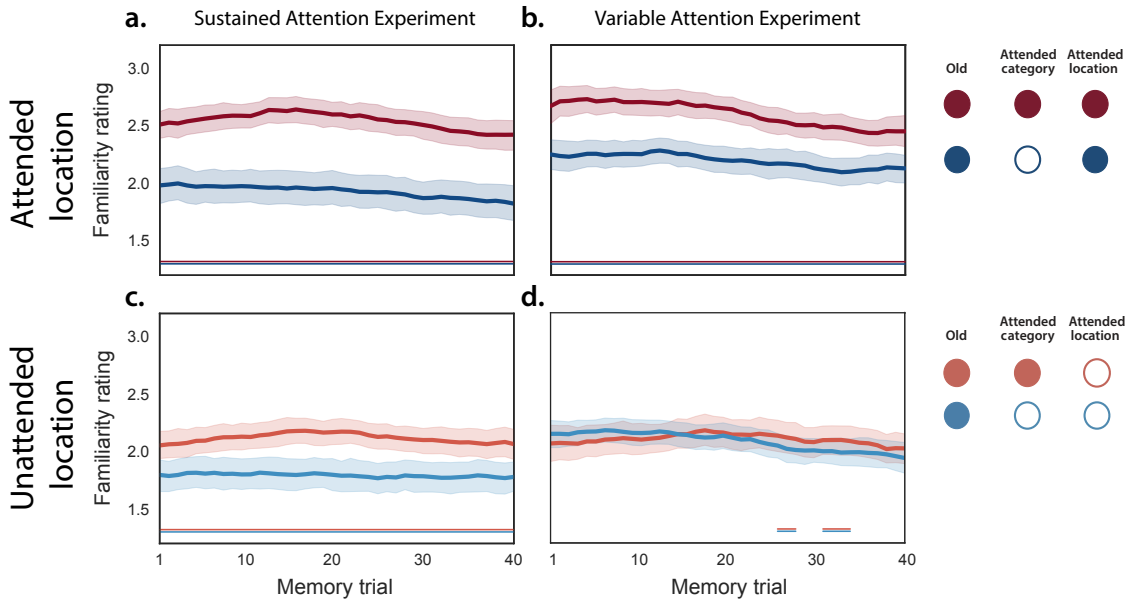


Figure 1.5: Familiarity ratings over time for images that matched the cued or uncued image location.

This figure follows the same format and conventions as Figure 1.4. Appendix A displays these results broken down by participant cohort.

versus Variable Attention Experiment participants was in how they rated unattended-category images. Participants in the Variable Attention Task rated unattended-category images as more familiar than did participants in the Sustained Attention Task; this pattern persisted throughout the memory phase and held for images at both the attended and unattended locations (compare heights of blue lines in the left versus right panels of Fig. 1.5). This suggests that lower familiarity ratings of unattended-category stimuli in the Sustained Attention Experiment are due to suppression of the encoding of unattended-category features (e.g., as opposed to enhancement of the encoding of attended-category features). This suppression effect appears to build over an interval that is longer than the duration of a single stimulus presentation (5.5 s).

Finally, we examined how participants in both experiments rated the familiarity of novel images. These ratings provide insights into how modulating the focus of attention to *prior* stimuli can affect the perception of *future* stimuli. We considered how participants rated

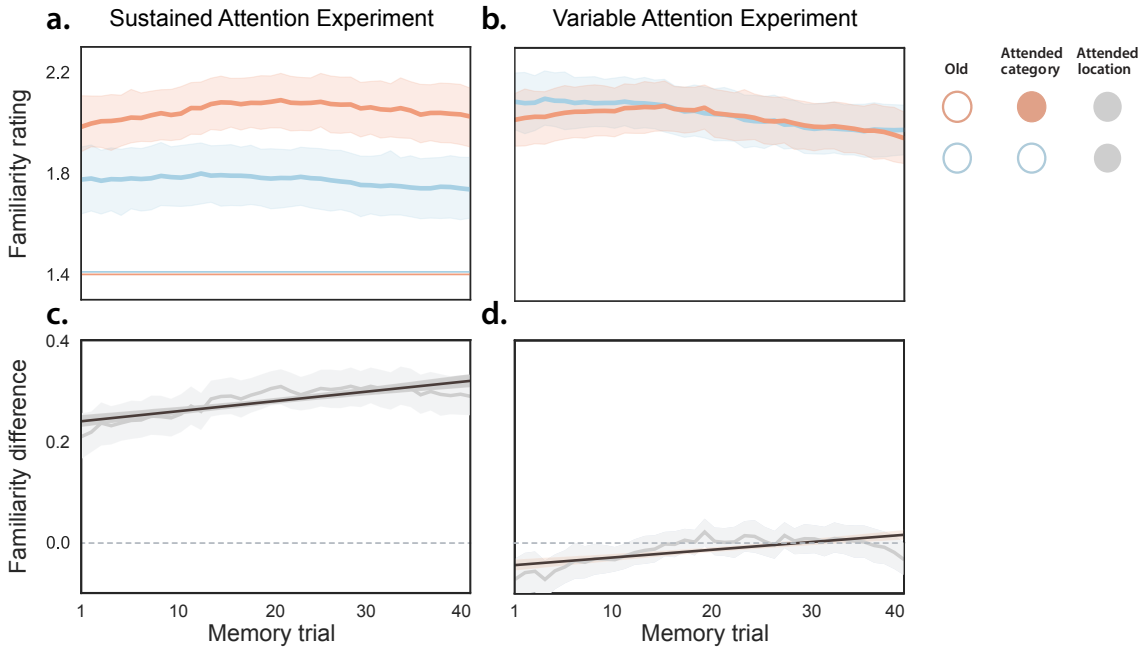


Figure 1.6: Familiarity ratings over time for attended-category and unattended-category novel images.

The top panels of this figure follow the same formats and conventions as Figures 1.4 and 1.5. The bottom panels display the differences between the timecourses of familiarity ratings of attended-category and unattended-category novel stimuli. The error ribbons denote 95% confidence intervals, computed across participants. The dotted horizontal line in each panel denotes a difference of 0– i.e., the value above which attended-category novel images were rated as more familiar than unattended-category novel images. The black lines in the bottom panels display linear regression fits to the data (error ribbons denote 95% confidence intervals). Appendix A displays these results broken down by participant cohort.

novel stimuli that matched versus did not match the category (face or scene) of the most recent attention cue. Participants in the Sustained Attention Experiment rated attended-category novel stimuli as more familiar than unattended-category novel stimuli throughout the entire memory phase (Fig. 1.6a). By contrast, participants in the Variable Attention Experiment rated attended-category and unattended-category novel images similarly throughout the memory phase (Fig. 1.6b). The difference in familiarity ratings between the two experiments primarily reflects that participants in the Variable Attention Experiment rated unattended-category novel images as more familiar than did participants in the Sustained Attention Experiment (e.g., compare blue lines in Fig. 1.6a versus 1.6b). This may indicate that the processing of unattended-category novel stimuli is suppressed following sustained attention (e.g., as opposed to enhancing the processing of attended-category novel stimuli). The differences in familiarity between attended-category versus unattended-category novel images increased slightly but reliably throughout the memory phases of both experiments (Fig. 1.6, bottom panels; slope of regression line fit to differences in Sustained Attention Experiment: $\beta = 2.05 \times 10^{-3}, p = 1.07 \times 10^{-14}$; slope of regression for Variable Attention Experiment: $\beta = 1.54 \times 10^{-3}, p = 5.63 \times 10^{-13}$). This might reflect a process whereby the suppression of unattended features continues to build for a short duration even after the participant stops volitionally focusing on the attended features.

1.4 Discussion

We ran two covert attention experiments that asked participants to sustain or vary the focus of their covert attention, respectively. When participants held the focus of their feature-based (face versus scene) and location-based (left versus right) attention for a sustained interval, they judged stimuli they had seen as familiar when they overlapped with respect to the features and locations they had attended. The increase in familiarity was

larger for attended-feature images than attended-location images. The increase also extended to novel stimuli from the attended image category. By contrast, when participants varied the focus of their feature-based and location-based attention more rapidly, the boost in familiarity for feature-matched stimuli was smaller than that for location-matched stimuli, and did not extend to novel stimuli. Our findings suggest that participants were able to more rapidly modulate their focus of location-based attention than their focus of feature-based attention. The tuning of location-based attention appears to be mediated by enhanced encoding and faster processing at the attended location. The tuning of feature-based attention appears to be mediated by a suppression in the encoding and processing of unattended stimulus features. This suppression effect also affects how new stimuli are processed, and it persists for a longer duration following an interval when the focus of feature-based attention was held constant over a longer duration. Taken together, our findings suggest that feature-based and location-based attention are mediated by different mechanisms and affect memory at different timescales (Fig. 1.7).

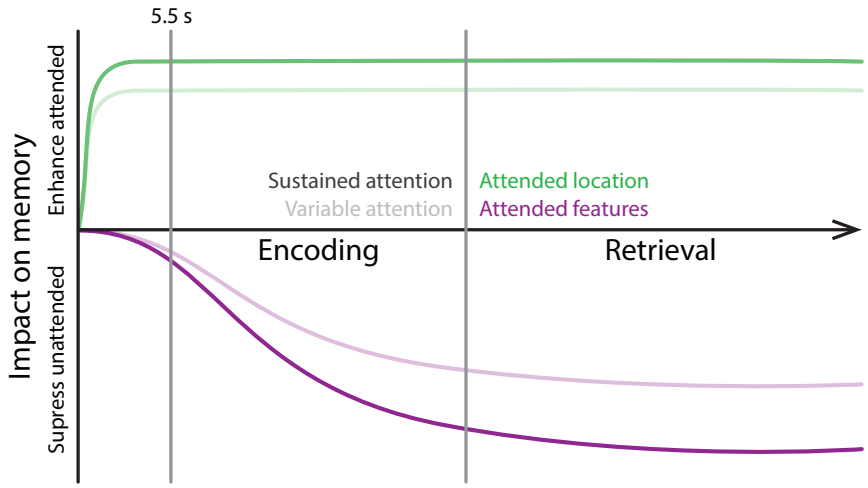


Figure 1.7: Timecourse of encoding and retrieval effects of feature-based and location-based attention.

The focus of location-based attention (green) may be modulated quickly and rapidly enhances memory encoding for nearby stimuli. The focus of feature-based attention (purple) is modulated more slowly, and serves to suppress unattended stimulus features. The effects of location-based attention on memory persist for longer than the effects of feature-based attention. Sustained attention (darker shading) yields more robust enhancement and suppression than shorter-term (variable) attention (light shading). The vertical gray line on the left denotes the duration of a single stimulus presentation (the upper bound by which location-based attention begins to affect encoding, and the lower-bound by which feature-based attention begins to affect encoding). The vertical gray line on the right separates encoding from immediate subsequent retrieval of the encoded information.

Chapter 2

Unexpected false feelings of familiarity about faces are associated with increased pupil dilations

Ziman K., Manning J. R. (2021). Unexpected false feelings of familiarity about faces are associated with increased pupil dilations. *Under revision, Psychonomic Bulletin & Review.*

2.1 Introduction

Imagine that you are observing a crowd of people when you suddenly and unexpectedly notice a childhood friend, whom you haven't seen in many years, milling amongst the group. You call out and wave, walking towards them. However, when you are able to get a better look, you realize that it isn't your friend at all– it's a stranger that you've never met before. Awkwardly, you withdraw your hand and pretend to melt back into the scenery.

Research on false memory has shown that people often "fill in" perceived gaps in their

recall by building on the scaffolding of their prior knowledge of the current context or situation (Deese 1959; Roediger McDermott 1995; Gallo 2006; Loftus 1997). In essence, we pattern complete missing information based on our expectations. But what leads us to mistakenly identify an *unexpected* novel face, place, object, experience, or situation as familiar? We hypothesized that some new insights might come from an unexpected data source: pupil dilations.

Our pupils constrict when we move from a dark setting into a bright one, and dilate when we move from bright to dark. This serves to protect our retina's photoreceptors in the presense of excessive light energy, and to increase the available light energy when it is more limited. However, this involuntary response is not solely related to the physical intensity or energy of the light shining on the retina. For example, similar pupil constrictions and dilations may also be observed in response to *perceived* brightness or darkness (e.g., in brightness illusions), suggesting that pupillary responses are in part driven by subjective experiences (Laeng Endestad 2012). Although brightness is perhaps the strongest driver of the pupillary response, a growing body of work has shown that pupil dilation also tracks with a wide variety of higher-order cognitive processes. For example, pupil dilations also reflect changes in affect and emotion (Oliva Anikin 2018; Siegle 2003), attention to high-level information (OE. Kang 2014), the focus of high-level attention (O. Kang Wheatley 2015), synchronization between individuals engaged in conversation (O. Kang Wheatley 2017), hormonal release (McCorry 2007), expected value or expected utility (Slooten 2018), surprise (Preuschoff 2011), and familiarity (Võ . 2007; Gardner 1974, 1975).

When pupillary responses reflect high-order cognitive processes, it can be difficult to specifically identify the underlying causes of those responses, in part because many of these high-order processes are inter-related or otherwise inter-dependent. For example, when we encounter something unfamiliar, it can be surprising; evoke a sense of curiosity, fear, joy, or another affective response; cause us to evaluate its expected utility; and so

on. Therefore, even well-studied and relatively stable pupillary response effects, such as the finding that our pupils dilate in response to familiar stimuli or experiences (Gardner 1974; Võ 2007; Heaver Hutton 2011; Goldinger Papesh 2012; Papesh 2012; Naber 2013; Kafkas Montaldi 2015; Mill 2016; Jacoby 1991; Mandler 1980; Godden Baddeley 1975; Vilberg Rugg 2008; Yonelinas 2002) can be difficult to interpret. The recognition memory processes that lead us to feel a sense of familiarity depend in turn on myriad factors and processes that are also associated with changes in pupil dilation (Faber 2017; Beukema 2019; Zekveld 2018; Kahneman Beatty 1966; Kahneman 1967; Ahern Beatty 1981; Fiedler Glöckner 2012; Einhäuser 2017).

Here we sought to tease apart the pupillary responses associated with the *feeling* of familiarity from those related to the recognition memory processes that enable us to recognize when a stimulus or experience is *truly* familiar. We designed two conditions of an eyetracking experiment that first asked participants to attend to a series of locations and stimulus features while unattended stimuli and features also appeared on the screen. The two conditions differed in whether the attention cues were consistent across a series of presentations (*Sustained Attention*) or whether they varied randomly with each stimulus presentation (*Variable Attention*). In both conditions, we then asked participants to perform a recognition memory task whereby they rated the "familiarity" of attended, unattended, and novel stimuli. We examined pupillary responses as participants attended different stimuli and as they later made their familiarity judgements. In addition to replicating several previously reported attention-related pupillary response patterns, we also report pupillary responses to *novel* stimuli (i.e., that participants had not seen before) that they nonetheless identified as familiar.

2.2 Materials and methods

We sought to determine if items that feel familiar elicit unique pupil dilation responses, even if they are not truly stored in memory. To answer this question, we leveraged our previously published data from an experiment designed to test the effects of attention on memory. The full dataset and analysis code for this manuscript are freely available to the public.

Experiment design

The experiment comprised a series of presentation blocks and memory blocks. Throughout the presentation and memory blocks, pupillometry data were collected using an Eyetribe eye-tracking system (Eye Tribe, The EyeTribe, Copenhagen, Denmark). Full experimental and methodological details may be found in (Ziman 2020).

Presentation blocks

During presentation blocks, participants viewed a series of composite image pairs (one on the left and one on the right of the screen) while keeping their gaze pointed towards a centrally located fixation cross. Each composite image comprised an equal blend of a contrast and brightness normalized grayscale image of a face and an outdoor scene. Participants also received a visual attention cue (Fig. 2.1a) prior to viewing the composite image pairs (Fig. 2.1b), directing them to attend to face or scene component (*category*) of the left or right image (*location*). The frequency with which the attention cue was changed varied across two experimental conditions: a *Sustained Attention* and a *Variable Attention* condition.

Sustained attention. In the Sustained Attention condition of the experiment ($n = 30$), participants received a single attention cue at the start of each presentation block. In other words, they kept their attention focused on the same image location and category

throughout all of the composite image pair presentations. The attention cues were organized across blocks such that location and category were counterbalanced over the course of the experiment.

Variable attention. In the Variable Attention condition of the experiment ($n = 23$), participants received a new attention cue prior to every image pair presentation in the presentation block. In other words, they varied the focus of their attention on an image-by-image basis throughout the duration of the presentation block. The location and category cues within and across blocks were counterbalanced over the course of the experiment.

Memory blocks

During memory blocks, participants were instructed (Fig. 2.1c) to rate how "familiar" each of a series of grayscale images seemed on a scale from 1–4. If participants felt unsure about how to respond, they were explicitly instructed to take their best guess. Each image participants judged (Fig. 2.1d) was drawn either from the set of grayscale face and scene images that they had studied (as part of a composite image pair) during the prior presentation block (*old* images), or from a separate set of images that the participants had not encountered before (*novel* images). The set of images judged in each memory block comprised half old images and half novel images. In turn, the set of old images comprised an equal mix of presented images that whose locations were versus were not attended, and whose categories were versus were not attended. Across all memory blocks, participants viewed (and rated) a total of 80 novel face images and 80 novel scene images.

Pupillometry data analysis and preparation

Our eyetracking system continuously sampled participants' eye gaze positions (mean accuracy: 0.5° visual angle; mean precision: 0.1° visual angle root mean squared error) and pupil diameters at 30 Hz. We excerpted three-second windows that began when each new

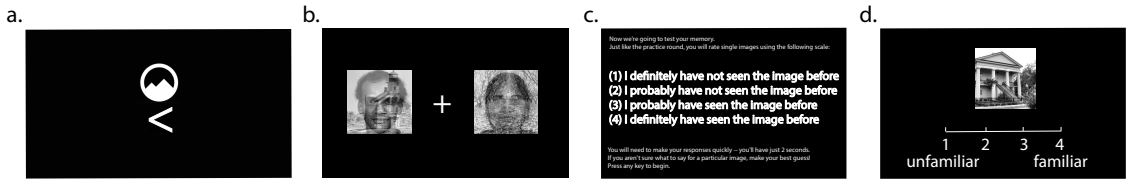


Figure 2.1: Covert attention & recognition memory paradigm.

a. During presentation blocks, participants received cues, like the one displayed here, directing their attention to the face or scene component of the left or right composite image. The example cue is directing the participant to attend to the scene component of the left image. **b.** An example composite image pair with a central fixation cross. **c.** Screenshot of instructions shown to participants prior to each memory block. **d.** An example image and familiarity rating response scale displayed during a memory block. Note that the scale of the text and images in all panels have been altered for illustrative purposes.

image or composite pair appeared on the participant's screen. For presentation trials, this window spanned the full duration that composite images displayed on the screen (3s). For memory trials, this window spanned the duration individual images appeared on the screen (2s) in addition to a fixation period after each image disappeared (1s).

We excluded samples where any of the following criteria held: the diameter for either pupil was measured as zero; the inter-pupillary difference in pupil diameters was greater than 1.5 times the interquartile range (across all trials); the gaze position was outside of the border of the display screen; the horizontal or vertical position was greater than 1.5 times the interquartile range from the average gaze location (across all trials); or the same sample was redundantly recorded. When the average sampling rate (for the remaining samples) dropped below 20 Hz, we removed those trials from our analyses. We estimated the pupil diameter (i.e., the pupil dilation response) at each timepoint by averaging the measured left and right pupil diameters. Finally, we converted these averaged diameters into z-scored (standard deviation) units within each participant.

We generated a smooth, regularly sampled, timecourse of the pupil dilation responses to each image by fitting a Piecewise Cubic Hermite Interpolating Polynomial (PCHIP;

Fritsch Carlson 1980) to the pupil diameters from each trial, and sampling from when the trial began until the moment the last viable pupil dilation measurement in the trial was recorded (rounded down to the nearest of 150 evenly spaced timepoints throughout this interval).

Segmenting the pupil response timecourse

The pupil response timecourses we observed during different parts of the experiment were often similar across participants. These timecourses often exhibited an initial dilation just after a new image appeared on the participant's screen, followed by a constriction, and so on. This suggested that different time intervals (relative to the image onset) might reflect different processes of potential interest. For each type of trial we analyzed (presentation trials, memory trial responses to old images, and memory trial responses to new images), we first computed the average pupil response timecourse across all trials and participants. We next computed the average value, m, of the average pupil response timecourse during the time interval of interest. We then segmented the pupil response timecourse into consecutive time bins where the pupil dilations were consistently above or below m. This yielded a set of "cut points" for the pupil response timecourse (i.e., mean crossings). Finally, we examined the pupil responses within each of these segments to identify potential differences in pupil dilations as a function of the familiarity ratings that the participants assigned (or would later assign) to different images.

2.3 Results

To explore pupillary responses under different attention and memory conditions, we computed the average pupil dilation response timecourses across trials and participants (see *Pupillometry data analysis and preparation*). We first examined pupillary responses as participants attended composite image pairs while keeping their gaze fixed on a central point

(Fig. 2.1b). We reasoned that these pupillary responses might reflect processes related to controlling the focus of feature-based or location-based (spatial) attention, or related to encoding the images into memory. We observed similar response timecourses across both experimental conditions (Sustained Attention, whereby participants were given the same cue for all composite image pairs within in a block; versus Variable Attention, whereby participants were given a new attention cue prior to viewing each image pair; see Fig. 2.1a for an example attention cue). Figure 2.2a displays results averaged across both experimental conditions; Figures and display analogous results broken down by condition. The average pupil dilation timecourse we observed when participants viewed the composite image pairs is displayed in Figure 2.2a. As summarized in Figure 2.2, participants' pupil dilation increased when they attended to images that they later recognized (i.e., rated as familiar during the memory phase of the experiment; blue curve in Panel b). When participants attended to images that they would later fail to recognize, their pupils did not dilate as much (red curve in Panel b). This suggests that participants' pupils were dilating when they successfully encoded an image from the attended location and category into memory. When we examined pupil responses to unattended images (i.e., images from the unattended location or category) we observed no reliable differences in participants' pupil response timecourses as a function of the familiarity ratings they assigned to those images during the memory phase of the experiment (Fig. 2.2c–e). This suggests that the unattended images may not have been encoded into memory as reliably, or that some other mechanism or process that does not track as closely with pupil dilations might govern the encoding of the unattended images.

Next, we examined pupillary responses as participants rated the familiarity of the images that had comprised the composite pairs they had seen during the presentation phase of the experiment. We reasoned that these pupillary responses might reflect processes related to memory retrieval. We observed similar timecourses across both the Sustained Attention and Variable Attention experimental conditions. Figure 2.3a displays results

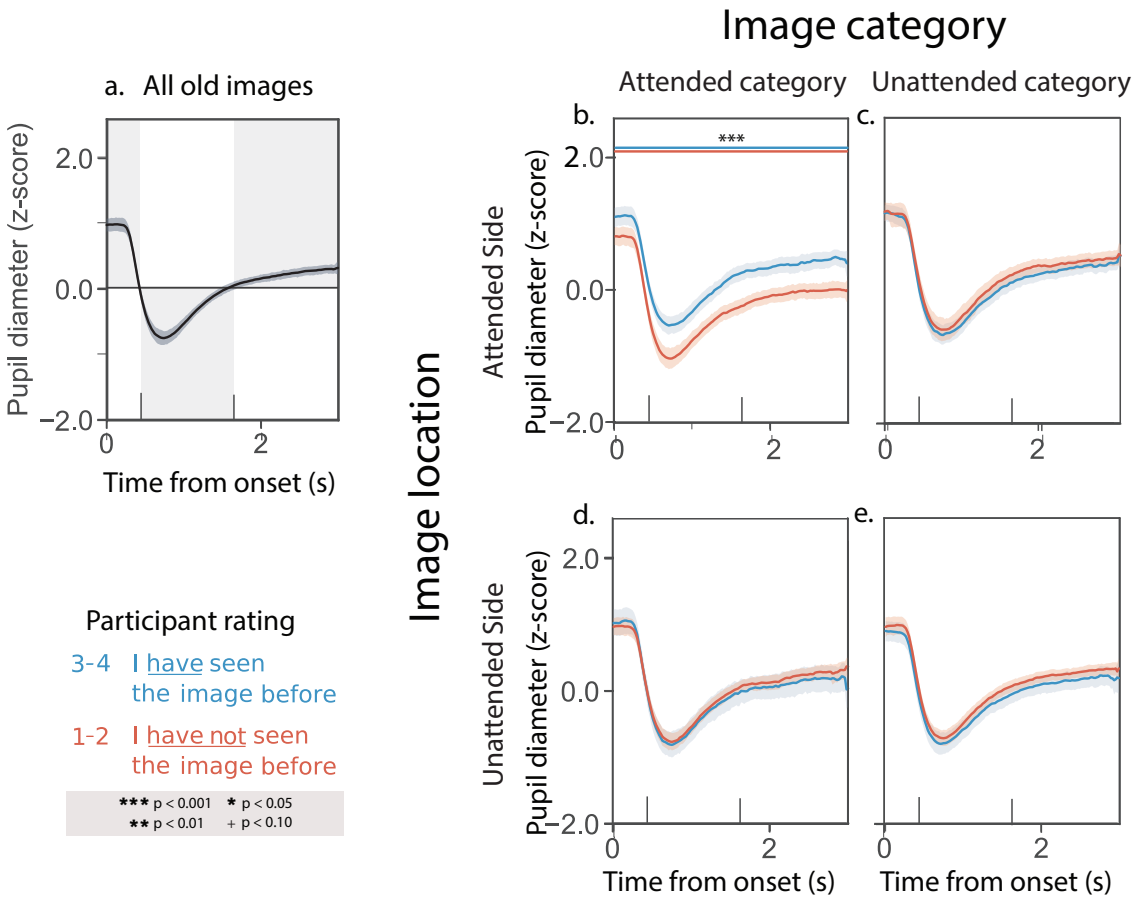


Figure 2.2: Pupil dilation response timecourses while attended to composite image pairs.

a. Average pupil dilation timecourse across all trials and experimental conditions. **b.** Pupil dilation timecourses for trials corresponding to attended images that participants later rated as familiar (blue curve; familiarity rating = 3 or 4) versus unfamiliar (red curve; familiarity rating = 1 or 2). **c.** Pupil dilation timecourses for trials corresponding to images on the attended side (but the unattended category) that participants later rated as familiar (blue) or unfamiliar (red). **d.** Pupil dilation timecourses for trials corresponding to images from the attended category (but on the unattended side) that participants later rated as familiar (blue) or unfamiliar (red). **e.** Pupil dilation timecourses for trials corresponding to images on the unattended side, from the unattended category, that participants later rated as familiar (blue) or unfamiliar (red). All panels: error ribbons denote 95% confidence intervals across participants. See Supplemental Figures and for analogous results broken down by experimental condition and numerical familiarity rating. In this figure, and in subsequent figures, the horizontal line pairs denote reliable separation (quantified using two-tailed paired *t*-tests) between the corresponding curves, during the time intervals covered by the lines. Significance levels are denoted by the symbols shown in the legend.

averaged across both experimental conditions; Figures and display analogous results broken down by condition. The average pupil dilation timecourse we observed when participants viewed previously seen memory cue images is displayed in Figure 2.3a. As summarized in Figure 2.3, participants' pupil dilation increased (numerically) when they recognized previously attended images as familiar (blue curve in Panel a) versus when they failed to recognize previously attended images as familiar (red curve in Panel a). We observed a qualitatively similar increase in pupil dilation when participants rated partially attended images as familiar (blue curves in Panels c and d) versus unfamiliar (red curves in Panels c and d). Although the responses displayed in Panels b–d are all qualitatively similar, only the differences in Panel d crossed our threshold for statistical significance. Finally, we saw no consistent familiarity-dependent changes in pupil responses to unattended images (Panel e).

Taken together, the above pattern of results could be consistent with several possible interpretations. One possibility is that participants' pupils dilate during memory retrieval, analogous to the responses we observed during the presentation phase of the experiment that appeared to track with memory encoding. This seems to be supported by the finding that differences in pupil dilation responses to images that were rated as familiar versus unfamiliar appear to fall off monotonically as a function of how much attention participants were instructed to pay to the corresponding images during the presentation phase of the experiment (e.g., compare Panels b–d with Panel e). In this way, our results thus far potentially agree with findings from myriad studies showing that people's pupils dilate when they are engaged in remembering or recognizing (Goldinger Papesh 2012; Haj . 2019; Rijn . 2012; Kucewicz . 2018; Naber . 2013; Mill . 2016). However, an alternative explanation is that pupil dilations might instead reflect the *feeling* of remembering or recognizing as opposed to memory retrieval per se. We hypothesized that participants' familiarity judgements of novel (never before seen) images might enable us to disentangle these explanations. In particular, if we observed a pupil dilation response during the

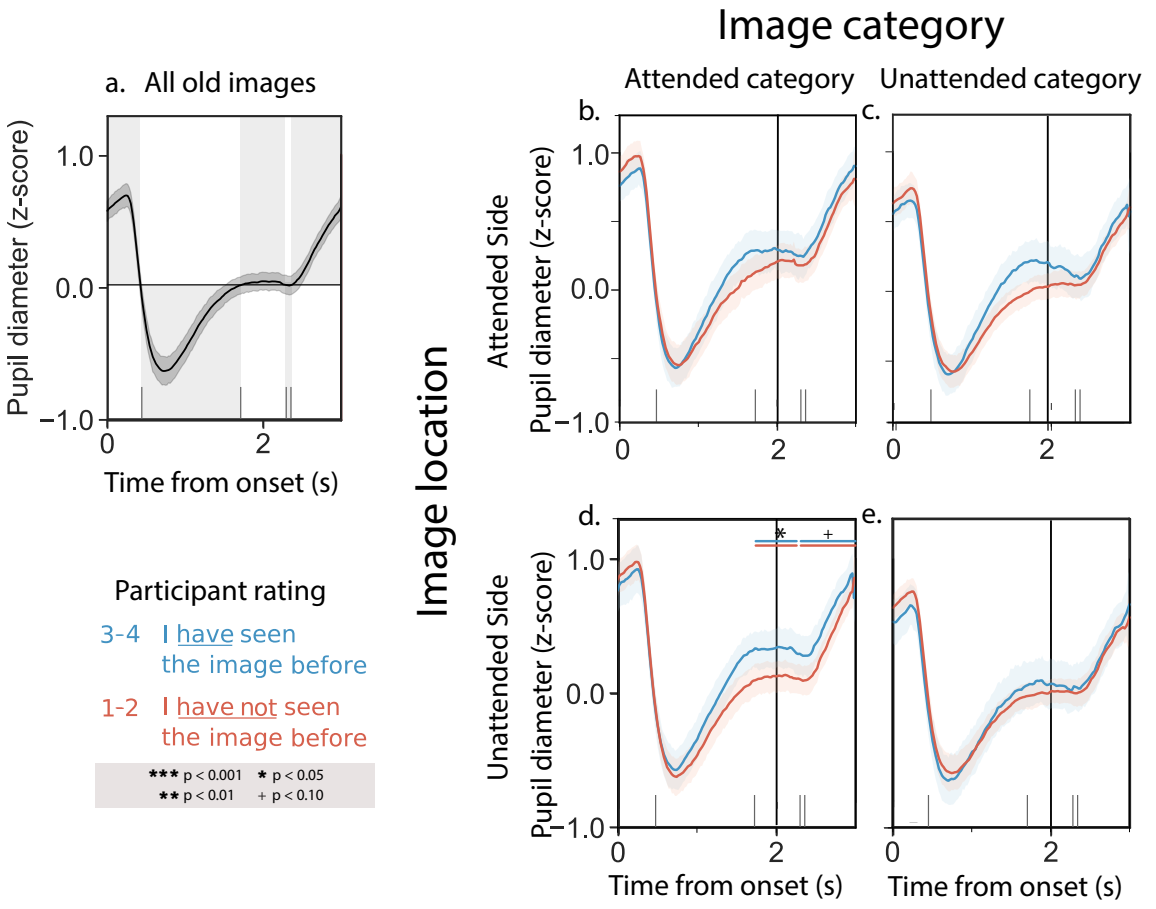


Figure 2.3: Pupil dilation response timecourses while rating the familiarities of previously studied images.

a. Average pupil dilation timecourse across all trials and experimental conditions. **b.** Pupil dilation timecourses for trials corresponding to previously attended images that participants rated as familiar (blue curve; familiarity rating = 3 or 4) versus unfamiliar (red curve; familiarity rating = 1 or 2). **c.** Pupil dilation timecourses during trials where participants rated images on the attended side (but the unattended category) as familiar (blue) or unfamiliar (red). **d.** Pupil dilation timecourses during trials where participants rated images from the attended category (but the unattended side) as familiar (blue) or unfamiliar (red). **e.** Pupil dilation timecourses during trials where participants rated unattended images as familiar (blue) or unfamiliar (red). All panels: error ribbons denote 95% confidence intervals across participants. The vertical lines indicate when the images were cleared from the screen. See Supplemental Figures and for analogous results broken down by experimental condition and numerical familiarity rating.

rare times when participants mistakenly rated novel images as familiar, this would indicate that the pupil response is driven in part by the feeling of familiarity rather than the specific engagement of memory retrieval processes.

When we examined participants' familiarity ratings of novel stimuli, we noticed several behavioral patterns. In the Sustained Attention condition, participants rated novel images as more familiar if they came from the most recently attended category (familiarity ratings of novel stimuli that matched versus conflicted with the most recent attention cue: $t(29) = 4.37, p < 0.001$). In the Variable Attention condition, participants tended to rate novel scene images as more familiar than face images, regardless of the most recent attention cue, although this tendency did not cross our threshold for statistical significance ($t(22) = 1.24, p = 0.23$). This suggests that when participants modulate their focus of attention to specific stimuli, the consequences to how they subsequently process images linger beyond the duration that the cues remain relevant. When attention cues were stable (i.e., in the Sustained Attention condition), participants responded in a biased way to novel images that matched the most recent stable cue. However, when the attention cues changed rapidly (i.e., in the Variable Attention condition), participants appeared to "default" to processing scenes and face images slightly differently, regardless of the most recent cued category. This suggests that different image categories may be processed or prioritized (in attention, memory, etc.) differently, independent of the specific experimental task, cues, or instructions. We therefore separated our further analyses of responses to novel stimuli along two dimensions: (1) whether or not the novel stimuli came from the most recently cued category and (2) whether the novel stimuli were scene versus face images.

Participants' pupillary responses to novel stimuli in the Sustained and Variable Attention conditions were similar. Figure 2.4 displays results averaged across both experimental conditions, and Figures and display analogous results broken down by condition. The average pupil dilation timecourse we observed when participants viewed novel memory

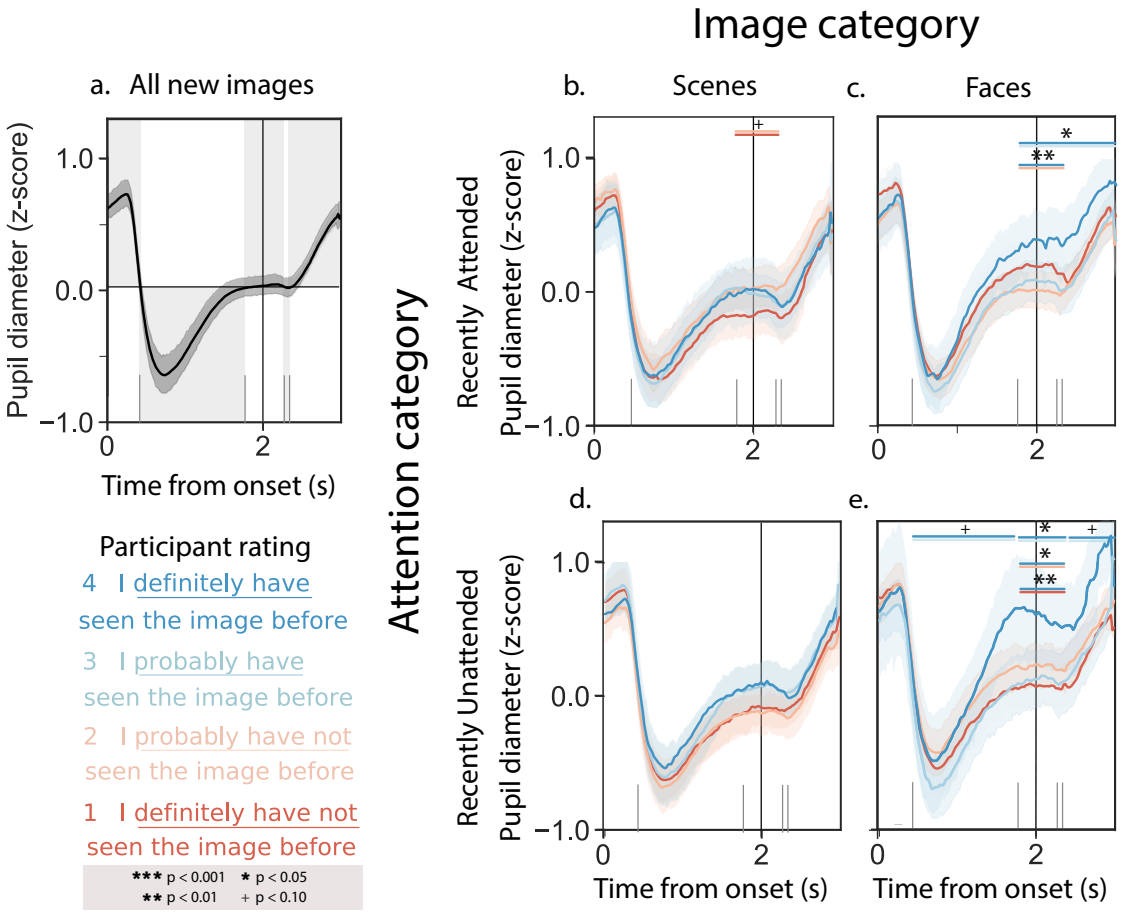


Figure 2.4: Pupil dilation response timecourses while rating the familiarities of novel images.

a. Average pupil dilation timecourse across all trials and experimental conditions. **b.** Pupil dilation timecourses (split by familiarity rating) for trials corresponding to novel scene images, when the most recent attention cue was also to a scene image. **c.** Pupil dilation timecourses (split by familiarity rating) for trials corresponding to novel face images, when the most recent attention cue was also to a face image. **d.** Pupil dilation timecourses (split by familiarity rating) for trials corresponding to novel scene images, when the most recent attention cue was to a face image. **e.** Pupil dilation timecourses (split by familiarity rating) for trials corresponding to novel face images, when the most recent attention cue was to a scene image. All panels: error ribbons denote 95% confidence intervals across participants. The vertical lines indicate when the images were cleared from the screen. See Supplemental Figures and for analogous results broken down by experimental condition.

cue images is displayed in Figure 2.4a. Unlike their responses to composite images during the presentation phase of the experiment, or to memory cues for previously seen images during the memory phase of the experiment, participants' pupillary responses to novel memory cues did not vary reliably as a function of the most recent attention cue (e.g., compare Fig. 2.4b versus d and c versus e). However, we did observe differences in participants' pupillary responses as a function of the category (scene versus face) of the novel memory cues. When participants viewed novel scene images, their pupillary responses showed no reliable differences as a function of the familiarity ratings participants assigned to those images (Fig. 2.4b and d). However, when participants viewed novel face images, their pupils dilated more when they rated the novel images as familiar (Fig. 2.4c and e).

2.4 Discussion

We examined pupillary responses as participants modulated their attention and rated the familiarity of previously seen and novel images. Whereas familiarity and retrieval are often conflated (e.g., when we recognize something we experienced in the past), examining pupillary responses to *novel* stimuli enabled us to disambiguate familiarity and retrieval. When participants rated novel faces as familiar, we observed a pupil dilation response that was qualitatively similar to the pupil dilation response we observed when participants correctly recognized previously encountered stimuli as familiar. However, the pupil dilation response to novel stimuli could not be explained by pure memory retrieval (since there were no prior memories about the stimuli to retrieve), nor could it be explained by response bias (since participants were biased to rate *scenes* as slightly more familiar than faces, all else being equal). Taken together, our findings suggest that the pupil dilation responses we observed are due to participants' *feelings* of familiarity. Further, this effect seemed specific to participants' responses to images of faces, in that we did not observe a familiarity-associated pupillary response when participants rated novel scene images.

We note several potential limitations of our study. The most substantial limitation we see is that we cannot entirely rule out that novel stimuli might trigger some sort of partial memory retrieval process. For example, a given novel image might *remind* a participant of other images they had encountered earlier on in the experiment. This could be driven by visual similarity, semantic similarity, or even associations drawn from the participants' prior experiences. This potential confound means that we cannot completely rule out that the pupil dilations we observed when participants rated novel faces as familiar might be driven in part by memory retrieval processes. However, any such process would need to be category selective, since we did not observe a pupil dilation response to novel scene images (regardless of their familiarity ratings). A second potential limitation of our study is that we cannot distinguish whether the pupil dilation response to familiar-seeming novel faces is specific to faces in particular, or whether it is instead category selective. To distinguish these possible explanations, one would need to collect additional data using images selected from a broader range of categories.

Our study contributes to a growing literature on pupillary responses in a wide range of cognitive tasks, particularly those aimed at studying processes underlying attention and memory (Korn Bach 2016). Prior work has also shown that our pupils dilate when we identify a target amidst a distracting background (Wang . 2020; Martin Johnson 2015), or when we detect an unexpected visual change (Kloosterman . 2015). Pupillary responses also track with internal belief states (Colizoli . 2018) and pre-conscious processes (Laeng . 2012). These findings help to contextualize our finding that participants' pupils dilated when they rated novel faces as familiar, even though they displayed an overall bias to rate novel faces as unfamiliar. The variety of cognitive phenomena that have been tied to pupillary responses also highlight the richness and complexity underlying the pupillary response. That a scalar value (pupil diameter) at a given moment incorporates such complexity also illustrates how difficult it can be to tease apart the many contributing factors. This also limits our ability to fully interpret pupillometry data (e.g., compared with

pure behavioral data, or some other biophysiological measurements under appropriate conditions).

The false feelings of familiarity our participants occasionally exhibited are also informed by a large literature on false memories (Deese 1959; Roediger McDermott 1995; Gallo 2006; Loftus 1997). Faces can be an especially interesting stimulus in these experiments given their special relevance and importance to everyday human life. Prior work on recognition memory for face images has shown that feelings of familiarity versus true memory retrieval-based recognition can be dissociated (e.g., by inverting the images Megreya Burton 2007), suggesting that these processes may be supported by different mechanisms. Other work has shown that familiarity can also be influenced by visual properties of the faces themselves (e.g., their visual distinctiveness Lewis 2010). Taken together, this work suggests that the feeling that something is familiar can be at least partially dissociated from remembering that something has been encountered before.

Chapter 3

Is automatic speech-to-text transcription ready for use in psychological experiments?

Ziman K., Heusser A. C., Fitzpatrick P. C., Field C. E., & Manning J. R. (2018). Is automatic speech-to-text transcription ready for use in psychological experiments? *Behavior Research Methods*.

3.1 Introduction

Speech-to-text engines became popular in the 1990s (Kurzweil . 1990) when the performance of speech recognition algorithms (primarily based on Hidden Markov Models; Rabiner 1989) reached sufficient levels to provide plausible, though still often inaccurate, transcripts (Bamberg . 1990). Recent advances in deep learning have ushered in a new era of substantially more accurate speech recognition (Hinton . 2012). Today, speech-to-text engines are ubiquitous, and are embedded into applications running on myriad devices ranging from phones to watches to thermostats to cars and beyond.

While automated speech decoding is now widespread in mainstream society, the technology has not yet been widely adopted by the psychological research community to facilitate analyses of verbal responses. However, speech decoding has the potential to save researchers an enormous amount of time when analyzing verbal response data, and to enable new experimental designs that adapt based on parameters derived from decoded speech data. Further, whatever their current limitations, as speech-to-text algorithms continue to mature, their utility in psychological research should improve as well.

We sought to explore the feasibility of embedding a modern speech-to-text translation engine into a psychological experiment that relies on verbal responses as its primary data source. As a proof of concept, we had participants study and verbally recall a series of random word lists. We had human annotators manually transcribe the recorded audio data (UPenn Computational Memory Lab 2015), and we also transcribed the data automatically using the Google Cloud Speech API (Halpern . 2016). We then carried out a series of analyses to compare the human-generated and computer-generated transcripts.

Overall, we found that the human-generated and computer-generated transcripts matched to a high degree. Our main interest was in assessing the extent to which the computer-generated transcripts recovered (with high fidelity, treating the human-generated transcripts as the "ground truth") the major patterns in free recall dynamics that have been well-reported in the literature Murdock (1962); Kahana (1996, 2012, 2017); Manning . (2015). We also identified points of disagreement between the two transcription methods, particularly in how they handled non-recall vocalizations. Our results suggest that automated speech-to-text transcription tools are mature enough to provide (within limits) a viable alternative to human annotation. This provides a potential means of carrying out verbal response experiments on thousands of participants on online platforms such as Amazon's Mechanical Turk (Crump . 2013). Furthermore, the possibility of incorporating this technology into experiments that adapt on-the-fly according to prior verbal

responses (where rapid ongoing manual transcription would be infeasible) is particularly exciting.

3.2 Methods

Participants

Thirty Dartmouth undergraduate students (22 female, aged 18–21) participated in our study. All participants had (by self-report) normal or corrected-to-normal vision, reading, memory, and attentional abilities. Each participant gave written, informed consent to volunteer for our study. They received course credit for their participation. Our experimental protocol was approved by the Committee for the Protection of Human Subjects at Dartmouth College.

Materials

We collected data in a sound-attenuated testing room, using a 27-inch 2016 iMac desktop computer. All audio was recorded using the iMac's built-in microphone. The experiment was implemented in jsPsych (de Leeuw 2015) and psiTurk (Gureckis . 2015), along with custom code for sending audio data to the Google Cloud Speech API.

Our stimulus set comprised a pool of 256 words chosen from an online repository of themed word lists (Col 2017). To create the word pool, we (manually) chose 8, 12, 16, or 20 common words from each of 15 semantic categories: body parts, building-related, cities, clothing, countries, flowers, fruits, insects, instruments, kitchen-related, mammals, states, tools, trees, and vegetables.

Our experiment code and data may be downloaded here. We also created an open-source Python toolbox for analyzing and plotting free recall data, and for automatically transcribing audio data (Heusser . 2017).

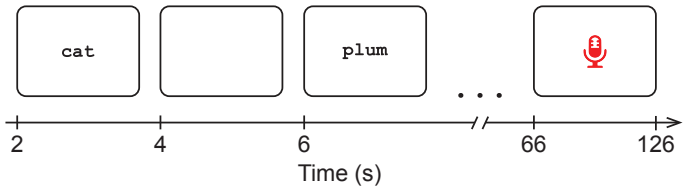


Figure 3.1: Free recall experiment paradigm

Experimental paradigm

Each participant studied a total of eight lists comprising 16 words each (128 words total). The lists were structured such that each contained four exemplars from each of four non-overlapping (but otherwise randomly selected from the pool) semantic categories, and each word appeared in (at most) one list. The specific set of 128 to-be-studied words were chosen anew for each participant, and the lists were generated randomly (with the words on each list shuffled randomly) for each participant. All text was displayed in black Courier New font, centered vertically and horizontally on a white background, and each letter was sized to occupy 5% of the screen width.

During each experimental *trial* (Fig. 3.1), the participant studied and recalled words from a single 16-word list. Each trial began with two seconds of blank white screen, followed by a two-second presentation of the first word on the list, followed by two more seconds of blank white screen. Each subsequent word was presented for two seconds, with a two second inter-stimulus interval (blank screen) before the next word's presentation. Two seconds after the last word was cleared from the screen, a red microphone icon appeared in the center of the screen, which prompted the participant to verbally recall as many words as they were able, from the just-presented list. The participant was given 60 seconds to recall the words "in the order they [came] to mind." Participants were instructed (at the beginning of the experiment) to speak "slowly and clearly" in order to facilitate analyses of their verbal response data. After 60 seconds, the microphone icon disappeared, the trial concluded, and the participant was given the opportunity to take a brief break before initiating the next trial.

Speech-to-text transcription

Each participant contributed a total of eight 60 s recordings of their verbal recalls of the studied word lists. We transcribed each recall recording into text manually using human transcribers (i.e., *human-generated*) and automatically using the Google Cloud Speech API (i.e., *computer-generated*).

Human-generated transcripts

Two co-authors of this paper (PCF and CEF) manually transcribed the audio data using a transcription software tool, Penn TotalRecall (UPenn Computational Memory Lab 2015). The transcribers listened to each trial's audio file in turn, played back at 1x speed, pausing or repeating playback as often as needed for them to be confident in their transcripts. Using the full 256-item word pool as a reference, any clear mispronunciations (e.g. "marimbo" instead of "marimba") or plurality errors (e.g. "hips" instead of "hip") were corrected to match the words in the word pool. In addition, any utterances that were judged by the transcribers to be non-recall vocalizations (e.g. "um," "wait, let me think...," etc.) were excluded from the transcript. These transcription decisions (to make pronunciation and plurality corrections, and to exclude non-recall vocalizations) were intended to highlight aspects of speech-to-text transcribing that human listeners might be especially well-suited to, relative to automated methods.

Computer-generated transcripts

We used the Google Cloud Speech API to produce a computer-generated text transcript of each participant's verbal responses. A total of 240 audio files, totaling four hours of recordings, were transcribed (eight one-minute recordings per participant, for each of the 30 participants). We passed the 256-item word pool to the automatic transcriber as a *speech context*, which provides "hints" to the speech recognizer about which words to expect. Note that we did not pass any information about which specific words were reflected

in any specific audio file, and only half of the total word pool was presented to any given participant. The speech recognizer returned, for each audio file, a list of automatically transcribed words and vocalization onset times. In addition, for each decoded utterance, the speech recognizer returned a confidence rating ranging from 0 (not confident) to 1 (highly confident); these confidence ratings roughly correspond to the estimated probability that the given word label matched the given speech utterance. The implementation details of the Google Cloud Speech API are proprietary, but the API is made publicly available here.

3.3 Results

We sought to evaluate the transcription accuracy of a modern speech-to-text engine applied to recordings of verbal responses from a list-learning experiment. We used the annotations of human transcribers as a benchmark. We carried out a preliminary analysis to assess the degree of absolute agreement between the human-generated and computer-generated transcripts. We then carried out a series of *post hoc* analyses to evaluate how well the computer-generated transcripts recovered the detailed recall dynamics (Kahana 2012, 2017; Manning 2015) reflected in the human-generated transcripts.

Transcription accuracy

An accurate computer-generated transcript should satisfy three basic criteria. First, it should have a high *hit rate*, in that the computer-generated transcript should contain each of the words also contained in the human-generated transcript. We defined the hit rate as the average (across lists) proportion of words in the human-generated transcripts that were also contained in the computer-generated transcripts. Second, it should have a low *false alarm rate*, in that the computer-generated transcript should *not* contain words that were not also contained in the human-generated transcript. We defined the false

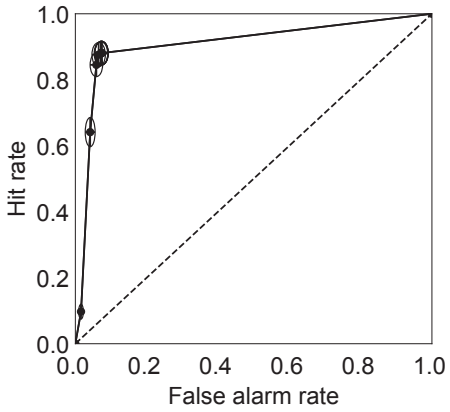


Figure 3.2: Receiver operating characteristic (ROC) curve for speech-to-text confidence ratings.

False alarm rate and hit rate as a function of the speech-to-text engine's confidence ratings (evaluated on the interval $[0, 1]$ in increments of 0.1). The ROC curve reflects an average across a total of 240 lists studied by 30 participants. Error ellipses denote 95% confidence intervals (across subjects).

alarm rate as the average (across lists) proportion of words in the computer-generated transcripts that were *not* contained in the human-generated transcripts. Third, for words in both sets of transcripts, the speech onset times should match well.

Because the speech-to-text engine we evaluated is probabilistic, each outputted response in the computer-generated transcripts is associated with a confidence rating. For this analysis, we used the receiver operating characteristic (ROC) to evaluate how the hit rate and false alarm rate varied as a function of the speech-to-text engine's confidence ratings (Fig. 3.2; area under the ROC: 0.907). Our analysis revealed that the human-generated transcripts matched the computer-generated transcripts well, in terms of the set of words each contained. This finding, that the two sets of transcripts matched well, indicates that the verbal responses transcribed by the speech-to-text engine were an accurate reflection of what the participants actually said (as judged by human observers). Note that in subsequent analyses we ignored the speech-to-text engine's confidence ratings (i.e., we included every transcribed word, regardless of rated confidence, in our analyses below).

In addition to evaluating the degree of match between the words identified in the human-

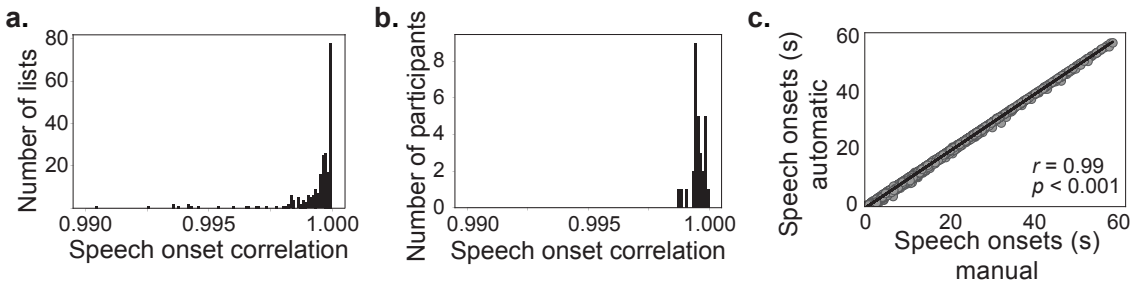


Figure 3.3: Speech onset times during recall.

a. Within-list correlations between human-generated and computer-generated speech onset times during recall. Each participant contributed data for eight lists. **b.** Within-subject correlations between human-generated and computer-generated onset times. **c.** Onset times for individual recalls, as identified manually and automatically. Each recall appears as a single dot.

generated and computer-generated transcripts, we also compared the speech onset times of words that appeared in both transcripts. We first correlated the onset times within list, whereby we obtained a total of eight correlations for each participant (one per list). The correlations on every list exceeded 0.99 (Fig. 3.3a). We next correlated the manually and automatically tagged onset times within subject, aggregating across all of the lists they encountered. We designed this analysis to catch potential failures of the speech-to-text engine to accurately identify differences in speech onset times across lists. Both sets of transcripts again displayed highly correlated onset times; all correlations exceeded 0.995 (Fig. 3.3b). Last, we correlated the manually and automatically tagged onset times of all recalls, aggregated across all lists and participants. We designed this analysis to catch potential failures of the speech-to-text engine to accurately identify differences in speech onset times across subjects. Again, the two sets of onsets times matched closely ($r = 0.99, p < 0.001$; Fig. 3.3c). Taken together, these onset time analyses indicate that the speech-to-text engine accurately identified speech onset times as identified by human annotators.

The above analyses show that, to a first approximation, the human-generated and computer-generated transcripts agreed well in terms of the words they contained and the times

at which those words were vocalized. We next sought to evaluate the degree to which the computer-generated transcripts captured the detailed recall dynamics reflected in the human-generated transcripts.

Recall dynamics

Participants in our free recall experiment studied and recalled random word lists. In general, the free recall literature has characterized participants' recall dynamics along four dimensions (for review see Kahana 2012). First, given a just-studied random word list, which word do participants tend to recall first? Second, in which order(s) do participants transition from recalling one word to the next? Third, which words do participants recall overall? And fourth, what sorts of errors (recalls of words that they had *not* studied) do participants make? We evaluated the degree to which the computer-generated transcripts captured each of these dimensions as compared with the human-generated transcripts.

Probability of first recall

The *probability of first recall* curves in Figure 3.4a display the proportion of trials in which participants began their recall sequences with words at each study position. In other words: which words on the just-studied lists did participants tend to recall first? The probability of first recall curves derived from the human-generated and computer-generated transcripts overlapped to a high degree. These curves indicate that participants in this experiment exhibited a strong *primacy effect*, whereby they most often began their recall sequence by recalling the first word presented on the just-studied list.

To better characterize the degree of agreement between the human-generated and computer-generated transcripts, and following our prior work (Manning . 2011), we defined a *primacy ratio* as the average probability of initiating recall with any of the first three words, divided by the average probability of initiating recall with any of the middle six words from the just-studied list. This yielded a pair of numbers for each participant; the first

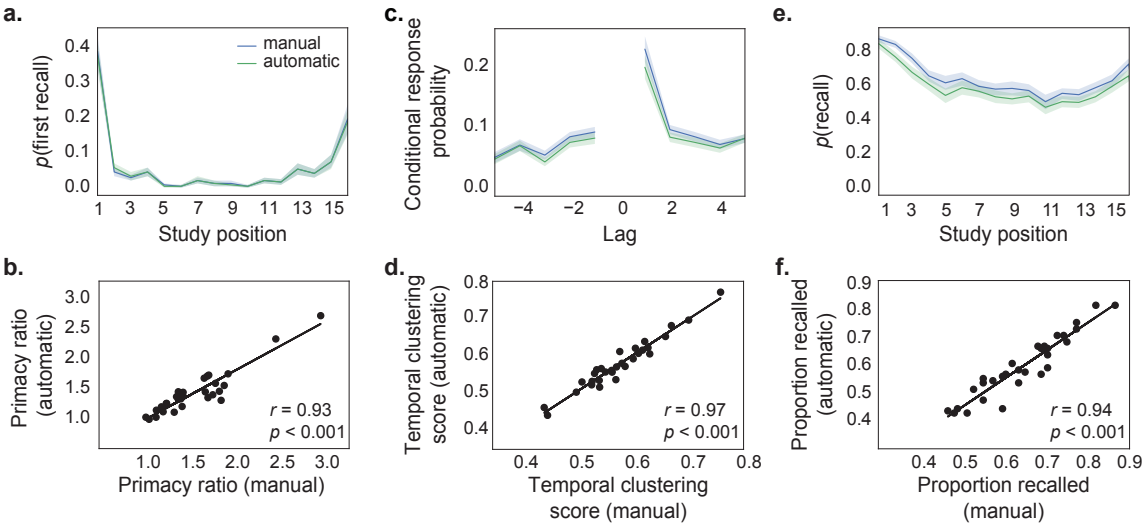


Figure 3.4: Recall dynamics during early, middle, and late recall.

a–b. Initiating the recall sequence. **a.** Probability of recalling each word first, as a function of its presentation position. Participants most often began their recall sequences with the first-presented word from the just-studied list. **b.** The "primacy ratio" (see text) reflects participants' tendency to initiate recall with words presented early (versus in the middle of the list). We assessed the agreement between the strength of this primacy effect as measured using the human-generated (*manual*) and computer-generated (*automatic*) transcripts. Each dot reflects the average primacy ratios for one participant. **c–d.** Recall transitions. **c.** The conditional probability of recalling each word as a function of its presentation position relative to the previously recalled word (lag). Participants often temporally cluster their recalls by successively recalling words that were presented at nearby positions on the list (Kahana 1996). **d.** We assessed the agreement between the degree of temporal clustering as measured manually and automatically. Each dot reflects the average temporal clustering scores (see text) for one participant. **e–f.** Overall recall probabilities. **e.** Probability of recalling each word as a function of its presentation position. **f.** We assessed the agreement between the average proportion of words recalled as measured manually and automatically. Error ribbons in Panels a, c, and e denote 95% confidence intervals (across subjects), estimated via 5,000 bootstrap iterations.

described the strength of the primacy effect as measured from the human-generated transcripts, and the second described the strength of the primacy effect as measured from the computer-generated transcripts. These two measures were highly correlated across participants ($r = 0.93, p < 0.001$; Fig. 3.4b), reflecting the high degree of agreement between the human-generated and computer-generated transcripts.

Recall transition probabilities

Given that a participant has just recalled a word from the just-studied list, which word are they likely to recall next? The lag Conditional Response Probability curves (Kahana 1996) displayed in Figure 3.4c reflect the conditional probability of recalling each word on the just-studied list as a function of its study position relative to the previously recalled word (*lag*). The curves show that participants tend to successively recall words that came from nearby study positions on the studied lists, a phenomenon referred to as *temporal clustering*.

Following Polyn Kahana (2008), we defined a *temporal clustering score* for each participant, reflecting their average tendency to successively recall words that came from nearby study positions. For each recall transition, we create a distribution of the absolute values of the differences (lags) between the study position of the just-recalled word and the set of words that had not yet been recalled. We then computed the percentile rank (in the distribution of absolute lags) of the next word the participant recalled. When we observed a tie, we assigned that recall the average percentile rank of all similarly ranked potential recalls. We defined the temporal clustering score as the average percentile rank across all recalls, from all lists, from that participant (we first averaged the ranks of recalls from each list, and then averaged across lists). If the participant always recalled the closest yet-to-be-recalled word, they would be assigned a temporal clustering score of 1. If the participant recalled the words in a random order (with respect to the words' study positions) this would yield a temporal clustering score of 0.5. We computed tempo-

ral clustering scores for each participant using both the human-generated transcripts and the computer-generated transcripts; the two transcripts yielded highly similar temporal clustering scores (Fig. 3.4d; $r = 0.97, p < 0.001$).

Overall recall probabilities

Prior work on free recall has established that participants are more likely to remember words that they studied at the beginning or end of a list, relative to middle words (these are often referred to as the *primacy effect* and *recency effect*, respectively; Murdock 1962). We plotted the proportion of words that participants recalled as a function of their study position and found that the human-generated and computer-generated transcripts agreed well and exhibited similar primacy and recency effects (Fig. 3.4e). We also considered the overall proportion of studied words that participants remembered, as measured using the human-generated and computer-generated transcripts. The two types of transcripts agreed well (Fig. 3.4f; $r = 0.94, p < 0.001$).

Taken together, the above analyses show that the specific words and onset times identified in the human-generated and computer-generated transcripts agreed well in terms of identifying the specific sequences of words participants remembered from the lists they studied, and the precise timing of each utterance. We next turn to a series of analyses aimed at characterizing the errors participants made, as identified using the human-generated versus computer-generated transcripts.

Recall errors

We first examined *prior list intrusion errors*, whereby participants mistakenly recalled a word from an earlier list in the experiment, rather than from the most recently studied list. Previous work has established that prior list intrusions are made more often from recently studied lists (e.g. the list before the most recent one) than from lists from much earlier in the experiment (e.g. five lists back in the experiment; for review see Kahana 2017).

Both the human-generated and computer-generated transcripts reflected this pattern, and agreed closely (Fig. 3.3a). For each participant, we also computed the average proportion of prior list intrusions that involved words from one list back, two lists back, and so on (up to five lists back). We computed these proportions using the human-generated and computer-generated transcripts and compared the results (Fig. 3.3b–f). We found that the numbers of prior list intrusions identified using both methods matched reliably (one back: $r = 0.95, p < 0.001$; two back: $r = 0.85, p < 0.001$; three back: $r = 0.99, p < 0.001$; four back: $r = 0.99, p < 0.001$; five back: $r = 0.74, p < 0.001$).

In addition to prior list intrusions, participants occasionally make *extra-list intrusion errors* by recalling words that had never been presented. Whereas the human transcribers intentionally filtered out non-recall vocalizations, the speech-to-text software effectively treated all vocalizations as recalls. For example, a human transcriber would treat the vocalization "the other two were also cities" as a non-recall vocalization, whereas the speech-to-text engine treats this as a series of six successive recalls. Similarly, a human transcriber would treat the vocalization "marimba, oh, did I already say harmonica?" as reflecting two recalls (of 'marimba' and 'harmonica'), whereas the speech-to-text software labels the utterance as a series of seven successive recalls. In other words, whereas human transcribers easily identify instances of participants "thinking aloud," automated methods do not distinguish recall from non-recall utterances.

Given this distinction between the manual and automatic transcriptions, we expected that there would (artificially) be a greater number of extra-list intrusions identified in the computer-generated, versus human-generated, transcripts (indeed, this pattern was reflected in our data; $t(29) = 7.15, p < 0.001$). However, although the computer-generated transcripts overestimated the numbers of extra-list intrusions, the numbers of extra-list intrusions identified in the human-generated and computer-generated transcripts were reliably correlated ($r = 0.61, p < 0.001$; Fig. 3.7). This indicates that the computer-generated

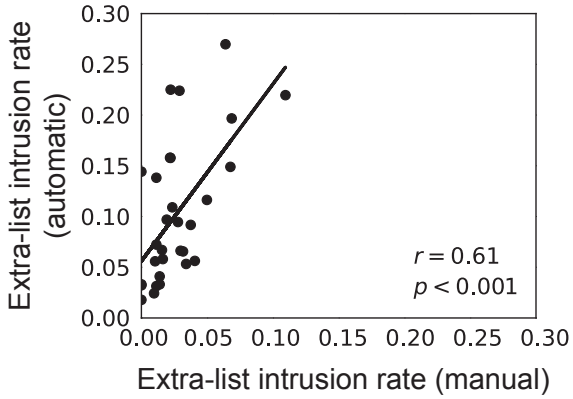


Figure 3.7: Extra-list intrusion errors.

Each dot reflects the average proportion of extra-list intrusions made by a single participant, as measured using the human-generated (x-axis) and computer-generated (y-axis) transcripts.

transcripts do not accurately reflect the *absolute* proportions of extra-list errors, but they do accurately reflect the *relative* proportions of extra-list errors.

3.4 Discussion

To gain insights into the extent to which modern speech-to-text engines might replace human annotators, we carried out a series of analyses on verbal responses recorded during a list-learning experiment. We found that the human-generated and computer-generated transcripts were largely in agreement. The computer-generated transcripts also accurately reflected most of the detailed statistical patterns that we identified in participants' recall behaviors. The major point of disagreement between the human- and computer-generated transcripts concerned how errors were reflected in the two types of transcripts. Whereas human annotators filtered out non-recall vocalizations, the computer-generated transcripts treated all vocalizations as recalls. This inflated the number of extra-list intrusion errors present in the computer-generated transcripts. Despite this, the computer-generated transcripts still accurately reflected individual variations in the relative num-

bers of errors generated by different participants. Overall, our results indicate that modern speech-to-text engines can accurately transcribe participants' verbal responses. To the extent to which direct transcripts are sufficient for capturing the behavioral phenomena of interest, our findings suggest that computer-generated transcripts can be used to capture and characterize verbal response patterns. This may carry substantial savings (of time and money) compared with human-generated transcripts. When large amounts of response data are collected (e.g. investigations into the effects of overt rehearsal on free and serial recall Rundus 1971; Tan Ward 2000, 2008) these savings may be particularly beneficial.

A note on our choice of speech-to-text engine

Our analyses in this manuscript leveraged a single speech-to-text engine (Halpern . 2016). We chose the Google Cloud Speech API due to its ease of use, the ability to provide a "speech context" (which played an important role in improving the transcription accuracy), the ability to obtain confidence ratings for each transcribed utterance, and the ability to automatically identify vocalization onset times. We have intentionally avoided detailed comparisons between this speech-to-text engine and the other promising speech-to-text engines available today that may have other advantages or disadvantages (e.g. Pocketsphinx; Huggins-Daines . 2006). Rather, the focus of our current analyses is to provide a proof-of-concept example of how modern speech-to-text engines can transcribe verbal response data. Our results highlight the immediate promise of existing speech-to-text technologies, and we expect that the quality of computer-generated transcripts will improve as the methods continue to mature.

Speech-to-text engines as a driver for scalable online verbal response experiments

Amazon's Mechanical Turk, launched in 2005, is an online marketplace that enables individual *requesters* to post small jobs that are carried out (usually in return for a small payment) by *workers* throughout the world. Over the past several years, psychological researchers and social scientists have begun to use Mechanical Turk as a convenient and low-cost platform for quickly collecting large amounts of experimental data (Paolacci . 2010). Despite the decreased level of control over the experimental environment relative to in-laboratory experiments, Mechanical Turk workers yield (for many, though not all, experiments) high quality behavioral data that are similarly reliable to data collected in the laboratory (Paolacci . 2010; Buhrmester . 2011; Crump . 2013). Recently developed tools like jsPsych (de Leeuw 2015) and psiTurk (Gureckis . 2015) facilitate the transition of laboratory experiments to the Mechanical Turk marketplace, substantially lowering the barriers to entry for research psychologists.

Our study suggests the feasibility of collecting verbal response data through Mechanical Turk. Whereas verbal responses are traditionally transcribed by human listeners (an approach that cannot easily scale to thousands of hours of recordings collected from thousands of participants via Mechanical Turk), automatic parsing via speech-to-text engines provides a potential avenue for quickly and cheaply transcribing vast quantities of data.

One potential downside to automated speech-to-text parsing is that these engines can be vulnerable to adversarial attacks, whereby malicious users intentionally generate stimuli that the speech-to-text engine will reliably transcribe incorrectly (e.g. Carlini Wagner 2018). While we would not expect this to be a widespread problem in typical online experimental settings, researchers are also beginning to devise strategies for counteracting adversarial examples (Madry . 2017). Nevertheless, adversarial examples (which a human observer would likely have transcribed correctly) provide another class of stimuli

that should be considered on an as-needed basis when applying these methods to massive online experiments that cannot easily be manually checked in detail.

Speech-to-text engines as a driver for adaptive verbal response experiments

Adaptive tests and experiments can dramatically reduce the time needed to assess knowledge and measure psychophysical and neuropsychological parameters. For example, computer adaptive testing is now widespread on standardized tests including the Graduate Record Examination (van der Linden Glas 2000), and the staircase method is commonly used to rapidly estimate participants' psychophysical thresholds (Cornsweet 1962). Modern variants of this technique, such as Bayesian active learning, use adaptive experiments to quickly map complex multivariate receptive fields based on neural data (Park Pillow 2012). Over the past several decades, researchers have also developed adaptive psychological experiments that leverage real-time processing of physiological signals, such as functional Magnetic Resonance Imaging (Cox . 1995; Cox Jesmanowicz 1999; Cohen 2001; deCharms 2008; deBettencourt . 2015) and electroencephalography (e.g. Angelakis . 2007).

Adaptive experiments driven by vocal responses have been limited, presumably because sufficiently accurate speech-to-text engines have only recently been broadly available. However, the computer-generated transcripts in our study captured many of the key patterns in participants' recall sequences. These transcripts may be generated on-the-fly during an experiment, for use in adapting future experimental trials (e.g. to optimize learning, more quickly converge on an estimate of participants' abilities or strategies, etc.).

Concluding remarks

The above findings show that automatic speech-to-text transcription, though imperfect, recovers many of the fundamental behavioral phenomena in free recall data. Our results

provide a "proof of concept" that automatic speech-to-text transcription is sufficiently accurate to serve as an effective substitute for human annotators in list-learning experiments. Additional study is needed to understand how broadly the level of performance we observed might generalize to other verbal response experiments, noisy recording environments, etc. Nevertheless, as improved speech-to-text algorithms are discovered and developed, we expect this to alleviate the need for human annotators.

Chapter 4

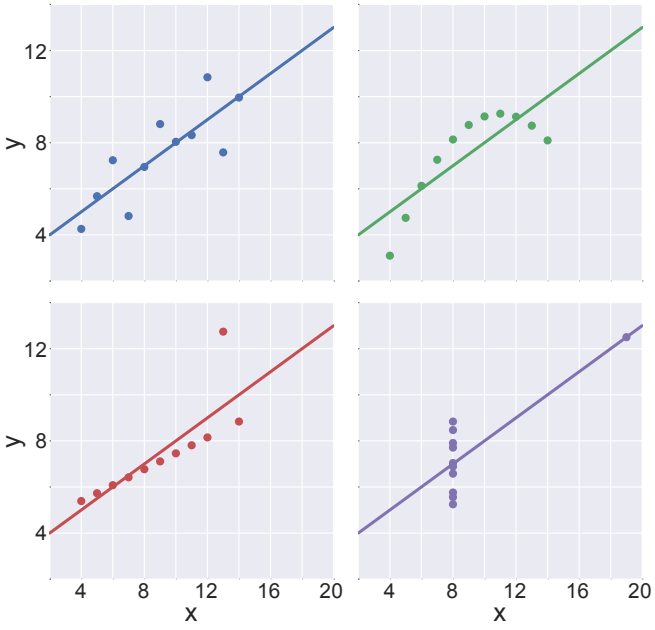
HyperTools: A Python toolbox for visualizing and manipulating high-dimensional data

Heusser A. C.*, **Ziman K.*** , Owen L. L. W., & Manning J. R. (2018). HyperTools: A Python toolbox for gaining geometric insights into high-dimensional data. *Journal of Machine Learning Research.*

4.1 Introduction

> "*To deal with hyper-planes in a fourteen dimensional space, visualize a 3D space and say 'fourteen' to yourself very loudly. Everyone does it.*"
>
> –Geoffrey Hinton (Hinton 2012)

The `HyperTools` toolbox is designed to reveal geometric structure in high-dimensional data through visualizations and manipulations. Modern data visualizations date back to at least the 16th century, when early data pioneers began to develop the sorts of accurate



	All Datasets
Mean (x)	9
Sample Variance (x)	11
Mean (y)	7.50
Sample Variance (y)	4.125
Correlation	0.816
Linear Regression	y = 3.00 + 0.500x

Figure 4.1: Anscombe's Quartet
Each dataset shares the same descriptive statistics but exhibits a unique shape.

maps and diagrams we might still recognize today (Friendly 2006; Tufte Graves-Morris 1983). Visualizations can reveal deep insights and intuitions about geometric structure and patterns in complex datasets by capitalizing on the human visual system's ability to quickly and efficiently extract meaning and structure from highly complex visual information (Uddenberg 2016). This is perhaps especially true of high-dimensional datasets, where different dimensions or features may interact in complex ways that may not be immediately obvious through conventional summary statistics.

As an illustration of the potential for summary statistics to mislead in the absence of visualization, consider the classic example, Anscombe's Quartet (Anscombe 1973) (Fig. 4.1). Anscombe's Quartet comprises four datasets that share a common statistical profile. Because the datasets are exactly equal along several common summary measures (mean, variance, trend lines), at first glance they seem highly similar. However, plotting the datasets and comparing them visually reveals that they differ substantially in structure. Whereas low-dimensional datasets like those in Anscombe's Quartet can be easily plot-

ted, it is not always obvious how to visualize high-dimensional datasets (e.g. with greater than 3 dimensions) in a similarly intuitive way.

A number of techniques, collectively referred to as *dimensionality reduction algorithms* have been developed over the past half-century to map high-dimensional data onto lower-dimensional representations that may be more easily manipulated and visualized. Some well-known examples include Principal Components Analysis (PCA) (Pearson 1901), Probabilistic Principal Components Analysis (PPCA) (Tipping Bishop 1999), Independent Components Analysis (ICA) (Jutten Herault 1991; Comon 1991), Multidimensional Scaling (MDS) (Torgerson 1958), and t-Distributed Stochastic Neighbor Embedding (t-SNE) (van der Maaten Hinton 2008). While the details of these algorithms differ, they each provide a means of obtaining a low-dimensional representation of the original high-dimensional dataset that preserves many of the geometric properties (e.g. the overall covariance structure of the data, data grouping, etc.) to the extent possible within a low-dimensional space. In the `HyperTools` toolbox, we leverage these dimensionality reduction algorithms (Fig. 4.2a) to aide in visualizing high-dimensional data.

A second class of algorithms leveraged in our toolbox provide techniques for manipulating and aligning different high-dimensional datasets (Fig. 4 .2b). These algorithms draw inspiration from the Procrustean transformation (Schönemann 1966), which computes the affine transformations (i.e. translation, reflection, rotation, and scaling) that bring one trajectory into alignment with another (in terms of minimizing the mean squared Euclidean distances between the corresponding points). The hyperalignment algorithm (Haxby . 2011) and the Shared Response Model (SRM) (Chen . 2015) extend this technique to find a common set of transformations that bring many (more than two) high-dimensional trajectories into common alignment. Our `HyperTools` toolbox leverages these alignment algorithms to allow users to manipulate and compare high-dimensional data, even when the dimensions (features) of original observations are different (e.g. brain patterns from

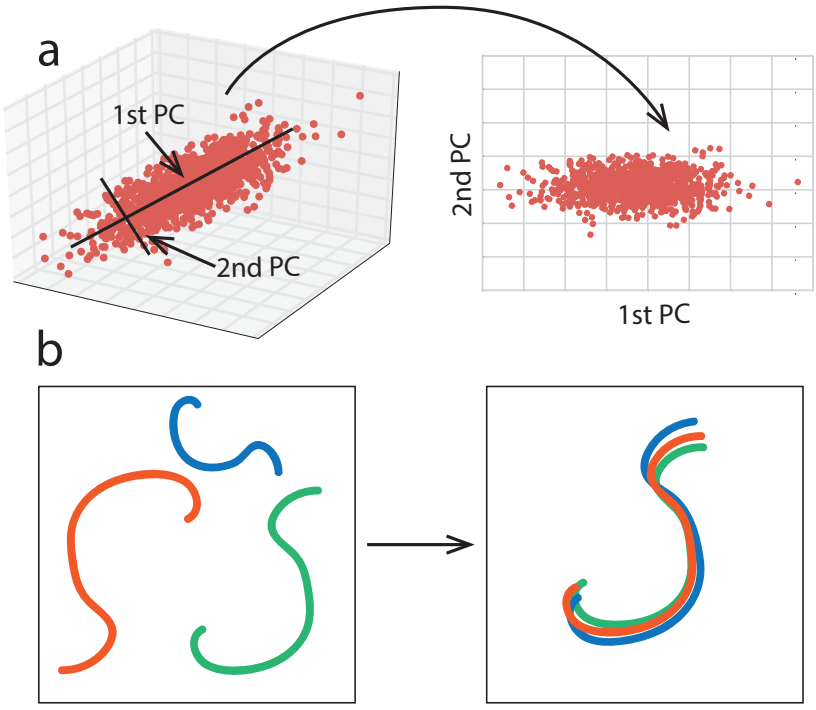


Figure 4.2: Visualizing and manipulating high-dimensional data

a. HyperTools uses dimensionality reduction algorithms to project high-dimensional data onto 2D and 3D plots. As shown in the panel, the dimensionality reduction algorithm PCA may be used to find the axes that explain the most variance in the original data (left panel). The data may then be projected onto a small number of those axes to facilitate plotting (right panel). **b.** Another important feature of `HyperTools` concerns aligning datasets with different fundamental coordinate systems. The left panel displays three trajectories with similar geometries but different coordinate systems, and the right panel displays how those trajectories may be aligned (via linear transformations) into a common space using hyperalignment.

different people, observations from different modalities, etc.).

Taken together, the `HyperTools` toolbox provides a set of powerful functions for visualizing and manipulating high-dimensional data using dimensionality reduction and data alignment algorithms. The toolbox is designed with ease of use as a primary goal, such that complex visualizations and analyses may often be carried out with a single line of code. Another major goal is to enable users to easily produce visually appealing publication-quality plots, also often with only a single line of code. Our toolbox is open-source and is distributed with the MIT License.

In the next section we provide a detailed overview of the components of the `HyperTools` toolbox and describe how the codebase is organized. We then describe a series of analyses of datasets from a wide array of domains to highlight many of the main functions of the toolbox.

4.2 Materials and Methods

Overview

The `HyperTools` toolbox is written in Python and can be downloaded from our GitHub page or with `pip`:

```
pip install hypertools
```
(4.1)

`HyperTools` depends on the following open-source software packages: `Matplotlib` (Hunter 2007) for plotting functionality, `Seaborn` (Waskom 2016) for plot styling, `scikit-learn` (Pedregosa, 2011) for data manipulation (dimensionatlity reduction, clustering, etc.), and `PPCA` for inferring missing data using PPCA. The toolbox also includes a port of the hyperalignment algorithm (Haxby 2011) from the `PyMVPA` library, as well as the shared

Filename	Description
`plot/plot.py`	Main plotting function: parses arguments, dispatches to `static.py` and `animate.py`
`plot/static.py`	Handles all static plot logic
`plot/animate.py`	Handles all animated plot logic
`tools/align.py`	Aligns the coordinate space of a list of matrices using hyperalignment
`tools/cluster.py`	Parcellates observations into discrete clusters using k-means clustering
`tools/describe_pca.py`	Analyzes and plots how many principal components are needed to capture the covariance structure of the data
`tools/missing_inds.py`	Find `nans` in data and returns indices
`tools/normalize.py`	z-scores rows/columns of matrices
`tools/df2mat.py`	Converts `Pandas` dataframes to `Numpy` arrays
`tools/procrustes.py`	Aligns the coordinate spaces of two arrays
`tools/reduce.py`	Reduces the dimensionality of one or more arrays using PCA and PPCA
`_externals/srm.py`	Implements the Shared Response Model (alternative alignment algorithm)
`_shared/helpers.py`	Collection of helper functions used across many files

Table 4.1
`HyperTools` Hypertools code organization.

The table lists the main files and functions that comprise the toolbox. We provide a feature complete description of the API on the project's GitHub page and in the documentation included with the toolbox download.

response model from the `BrainIAK` toolbox, as an alternative data alignment technique. In addition to providing a simple interface to several functions from these libraries, `HyperTools` adds a number of custom arguments to facilitate data visualization and manipulation of high-dimensional data. Table 4.1 provides summary of the `HyperTools` code base. In the remainder of this section, we provide descriptions of the primary toolbox functions, but we have not provided an exhaustive list. A feature-complete description of the API may be found on the project's GitHub page and in the documentation included with the toolbox download.

Nearly all of the `HyperTools` functions may be accessed through the main `plot` function. This design enables complex data analysis, data manipulation, and plotting to be

carried out in a single function call. There are two general types of plots supported by the toolbox: *static plots* and *animated plots*.

Static Plots

Accessing the `HyperTools` plot functionality entails first loading the to-be-analyzed dataset into the Python workspace and converting it to a `Numpy` array (van der Walt . 2011) or a `Pandas` dataframe (McKinney 2010). The format of the data should be samples (S) by features (F). Once the dataset conforms to this format, simply import the library and call the plot function:

```
import hypertools as hyp
```
(4.2)

```
hyp.plot(data)
```
(4.3)

By default (i.e. with no additional arguments specified), this function will perform dimensionality reduction (using PCA), convert the $S \times F$ `data` matrix into an $S \times 3$ matrix, and then create an interactive 3D line plot that can be explored by using the mouse to rotate the plot. If there are `nans` present in the dataset, these missing values will be automatically interpolated using PPCA (Tipping Bishop 1999). If $F < 3$, a 2D plot is created instead of a 3D plot. This simple interface to plotting is deceptively powerful: with a single command, the toolbox automatically fills in missing data and determines whether to create a 2D or 3D plot (reducing the dimensionality of the observations as needed).

`HyperTools` can also accommodate lists of `Numpy` arrays or `Pandas` dataframes (only single-level indexed dataframes are currently supported):

```
hyp.plot([array1, array2, array3])
```
(4.4)

When passed a list of arrays, `HyperTools` will plot each array in a distinct color. Colors

and styling can be customized in a several ways. Like `Matplotlib`, `HyperTools` can parse format strings passed as positional arguments. For example:

```
hyp.plot(array1, 'k-')
```
(4.5)

```
hyp.plot([array1, array2, array3], ['bo', 'r-', 'g:'])
```
(4.6)

Line colors may also be specified via the `color` (or `colors`) keyword argument:

```
hyp.plot(array1, color='g')
```
(4.7)

```
hyp.plot([array1, array2, array3],
         colors=['b', '#FF0000', (.3, .5, .4)]
```
(4.8)

Colors may be defined using format strings, hex codes, RGB values, or a mix of these formats. Rather than specifying the colors of each data array, colors may instead be specified for each individual sample by providing labels for each sample:

```
hyp.plot(data, group=group_labels),
```
(4.9)

where `group_labels` is a list of length S (number of samples). (Lists of group label arrays are also supported, e.g. if the data are passed in as a list of arrays; the list of labels must be of the same length as the list of data arrays.)

`HyperTools` parses this function call by sub-dividing each data matrix into new lists defined by each unique label in `group_labels`. For example, if each sample label is a string from the set `('a', 'b', 'c')` then each of these unique labels will be assigned a unique color, and the datapoints assigned to each label will be assigned that label's color.

In addition to specifying string labels for each sample, HyperTools also supports numerical labeling. If group_labels is a list of numbers, HyperTools will bin the range covered by those numerical values (excluding nans, Nones, and infs) into n evenly spaced bins (default: $n = 100$) and map these values onto a color palette. The color palette used for this mapping may also be customized using the palette keyword:

```
hyp.plot([array1, array2], palette='husl')
```
(4.10)

All Matplotlib and Seaborn color palettes and plot styles are supported by HyperTools.

In addition to specifying group-level labels (which are used to determine the colors of each sample), each sample may also be labeled with an additional text label that may be shown (as text) on the plot. The label and labels keyword arguments allow the user to define a list of strings (or a list of lists of strings) to be displayed next to each sample datapoint with an arrow pointing to it. Each list of labels must be of length S (number of samples). (The None value may be used to specify "blank" labels, which will not show up on the plot.)

By default, all datapoint labels are shown if the label or labels keyword is specified. However, HyperTools also supports a "data exploration" mode whereby the datapoint labels will only be shown when the mouse pointer hovers over the corresponding datapoint:

```
hyp.plot(data, labels=['a', None, 'a', 'b'], explore=True)
```
(4.11)

This plotting mode is especially useful when there are many datapoints, or when the data labels are long. If explore is set to True and no labels are specified, the labels will be auto-generated as an index and the PCA coordinate (e.g. '45: (3.0, 4.0, 5.0)'). Note: at the time of this writing, the labels and explore arguments are

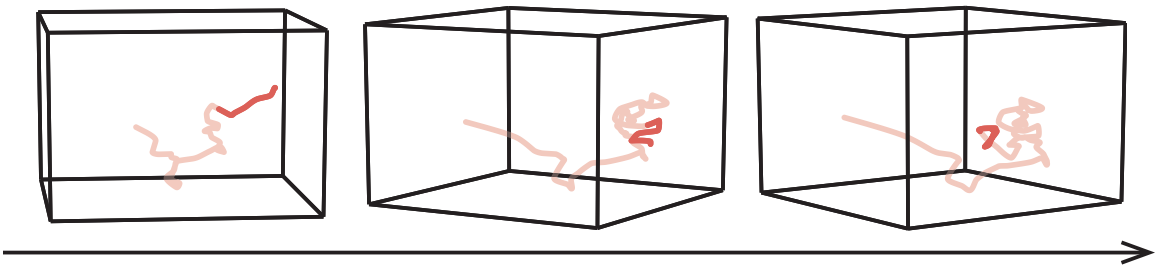


Figure 4.3: Frames from an animated plot.

Three frames (with time increasing moving from left to right) from an example animation are displayed in each panel.

only supported for 3D static plots.

Animated Plots

Animated 3D plots are especially useful for visualizing high-dimensional timeseries data.

To create an animation, simply toggle the `animate` keyword:

```
hyp.plot(data, animate=True)
```
(4.12)

This will create a 3D animated representation of the data, where the animation occurs over the rows of the `data` matrix. As with static plots, the user may pass a list of data matrices to plot multiple datasets on a single plot, and format strings and keyword arguments may be used to customize the plot appearance. Each frame of the animation displays a portion of the total data trajectory enclosed in a cube (Fig. 4.3). In successive frames, the displayed portion of the data trajectory advances by a small amount, and the camera angle rotates around the cube, providing visual access to different aspects of the data as the animation progresses.

The formats of animated `HyperTools` plots may be customized using the following keyword arguments: `duration` specifies the animation duration in seconds, `tail_duration` specifies the duration of the trailing tail in seconds, `rotations` specifies the number

of camera rotations around the data (over the course of the entire animation), `zoom` will zoom the camera in (positive number) or out (negative number) from the data, and setting `chemtrails=True` will plot a low opacity version of samples prior to the currently active trajectory so that the full structure and history of the data may be visualized. For a complete list of animation-specific arguments, please see the API documentation. Both animated and static plots can be saved by including the `save_path` argument (with the file extension included):

```
hyp.plot(data, save_path='path/to/the/file.pdf')
```
(4.13)

```
hyp.plot(data, animate=True, save_path='path/to/the/file.mp4'),
```
(4.14)

where Snippet 4.13 saves a static plot to a resolution-independent PDF, and Snippet 4.14 saves an animation as an MP4 movie. Note that saving animated plots requires that `ffmpeg` is installed on your computer; see the API documentation for more details.

Reduce

When passed high-dimensional data, the `plot` function uses PPCA to fill in missing data and PCA to project the data onto 3 dimensions. We provide API access to the `reduce` function that underlies these transformations. At its core, the `reduce` function is primarily a wrapper for `scikit-learn`'s PCA function and the PPCA package. `HyperTools` extends the functionality of these tools by providing an easier-to-use syntax and adding support for lists of matrices. The function may be used as follows:

```
reduced_data = hyp.tools.reduce(data, ndims=3)
```
(4.15)

Because dimensionality reduction results in information loss (relative to the original dataset),

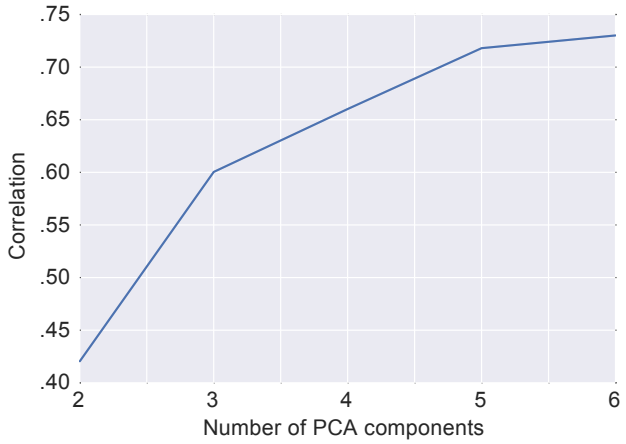


Figure 4.4: Covariance preserved as a function of the number of principal components.

For an example dataset (Example 2, *Results*) the panel displays the correlation between the upper triangle of the across-sample covariance matrices of the reduced versus original data, as a function of the number of principal components.

it is important to consider how accurately a low-dimensional projection of the data reflects the original high-dimensional dataset. `HyperTools` includes a function that plots the correlation between the covariance matrices of the reduced and full datasets, as function of the number of principal components in the reduced dataset (Fig. 4.4):

```
fig, ax, data = hyp.tools.describe_pca(data)
```
(4.16)

The `describe_pca` function computes these correlations iteratively (i.e. starting with one principal component, then two, then three, etc.) until a local maximum is detected. The resulting plot provides insights into the increase in explanatory power (in terms of the across-sample covariance) associated with each new principal component.

Align

Two or more datasets may share geometrical structure, but reside in different coordinate systems. Hyperalignment is a method that aligns the representational spaces over a list

of datasets, effectively co-registering them to a common space (Haxby . 2011). Using linear transformations, hyperalignment find a common space that minimizes the distance between two or more datasets (Fig. 4.2b). Aligning them to a common space allows one to visualize commonalities between the two different kinds of data. The `align` function accepts a list of arrays as input and returns a hyperaligned list of arrays in a common geometric space:

```
hyperaligned_list = hyp.tools.align([array1, array2, array3])
```
(4.17)

In addition to supporting alignment via the hyperalignment algorithm proposed by (Haxby 2011), we have also added support for alignment via the Shared Response Model (Chen 2015), which was ported from the `BrainIAK` toolbox:

```
SRM_aligned_list = hyp.tools.align([array1, array2, array3],
method='SRM')
```
(4.18)

Cluster

Some datasets exhibit *clustering* tendencies, whereby the data may be divided into discrete groups of similar or related samples (i.e. samples that are comprised of similar features). When these discrete groups are unlabeled or unknown, clustering algorithms provide heuristics for recovering these clusters of similar samples automatically. `HyperTools` incorporates the k-means clustering algorithm (Hartigan Wong 1979) to facilitate automatic data clustering. Given a pre-chosen number of clusters, k, the `cluster` keyword argument to the `plot` function uses k-means clustering to automatically assign each observation to a cluster, and then colors each observation's point according to its cluster

membership:

```
hyp.plot(data, n_clusters=k)
```
(4.19)

We also expose the k-means clustering algorithm directly through the `cluster` function:

```
cluster_labels = hyp.tools.cluster(data, n_clusters=k)
```
(4.20)

The `cluster` function wraps the `scikit-learn` implementation of k-means clustering and extends it to work with lists of data matrices.

4.3 Results

Example 0: Visualizing hypercubes in 3D

To illustrate how a user might visualize high-dimensional data with `HyperTools`, we start by examining four synthetic datasets with unique, known structures. We generated datasets of one cube (3 dimensions) and three hypercubes of increasing dimensionality (4, 5 and 6 dimensions), each comprised of 100 points along each of their respective edges. We then used `HyperTools` to project the hypercubes into 3 dimensional space (using PCA) and visualize the result.

Figure 4.5 illustrates how projecting hypercubes of different dimensionalities into 3 dimensional space distorts some aspects of their shapes, while preserving others. In the original (high-dimensional) data, all edges of each respective cube are of equal length, and each vertex comprises n adjacent edges converging orthogonally (where n is the dimensionality of the hypercube). However, in Fig. 4.5, some edges appear longer than

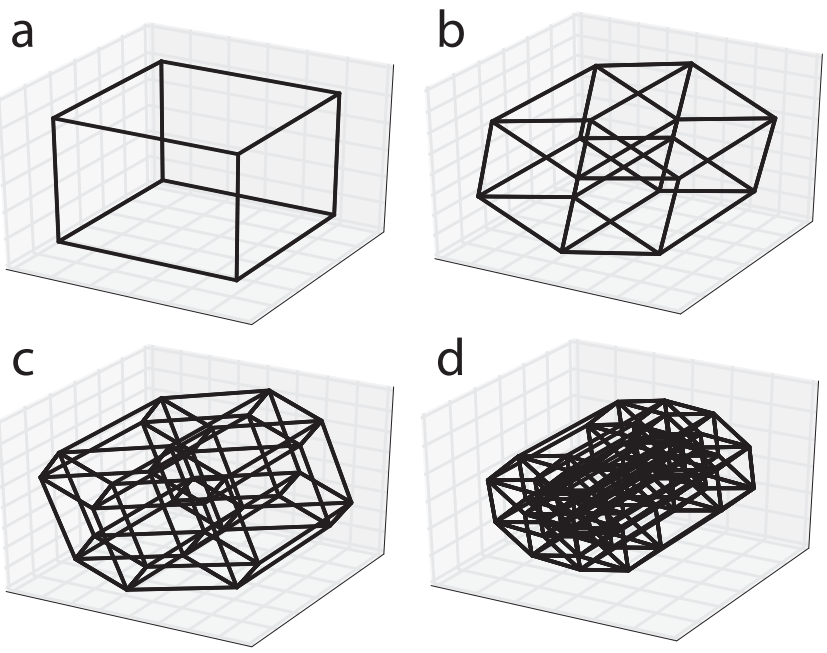


Figure 4.5: Hypercubes with increasing dimensionality.

Each dataset comprises 100 evenly spaced points along each edge of the corresponding cube with dimensionality **a.** 3, **b.** 4, **c.** 5, and **d.** 6.

others, and some vertices appear to form acute and obtuse angles.

Despite these differences, many of the underlying structural components are accurately reflected in the visualization. Namely, the visualization of each n-dimensional cube correctly depict 2^n vertices, $2^{n-1} * n$ edges, and n edges converging at each vertex. Each edge is also reliably reconstructed as a straight (rather than curved) line segment. The visualizations also depict increasing complexity with increasing dimensionality.

Example 0: Dimensionality reduction and clustering with various types of mushrooms

In this section, we highlight the dimensionality reduction and clustering capabilities of `HyperTools`. We retrieved the 'mushroom classification' dataset from the Kaggle database. The dataset contains annotated descriptive features of 8,124 mushrooms spanning 23

	class	cap-shape	cap-surface	cap-color	bruises	odor	...	habitat
0	p	x	s	n	t	p	...	u
1	e	x	s	y	t	a	...	g
2	e	b	s	w	t	l	...	m
3	p	x	y	w	t	p	...	u
4	e	x	s	g	f	n	...	g

Table 4.2
Example of mushrooms dataset.
The dataset contains annotated features (columns) of each mushroom (row), along with labels indicating whether each mushroom is poisonous or non-poisonous (not shown).

mushroom species from the *Audubon Society Field Guide to North American Mushrooms* (Lincoff National Audubon Society 1981). Each observation comprises a list of 22 descriptive features (e.g. cap shape, cap surface, habitat, etc.) along with a tag identifying each mushroom exemplar as poisonous or non-poisonous (features for five example mushrooms are shown in Tab. 4.2).

Because the mushroom features are provided as character strings, they must be transformed into numerical vectors to plot them. When passed a `Pandas` dataframe with columns containing text, `HyperTools` automatically converts the data into a binary matrix, where each column reflects one of the unique values of one of the features. The underlying function for converting dataframes into matrices may also be called directly:

```
matrix = hyp.tools.df2mat(dataframe)
```
(4.21)

Plotting the resulting matrix with `HyperTools` reveals a striking clustered structure. Overall, the samples appear to cluster by whether or not they are poisonous, but they also appear to group into sub-clusters (Fig. 4.6). By default, `HyperTools` uses PCA for dimensionality reduction, but different dimensionality reduction techniques can reveal distinct geometrical properties of a dataset. To highlight this, we plotted the transformed binary matrix using several dimensionality reduction techniques (PCA, ICA, t-SNE, and

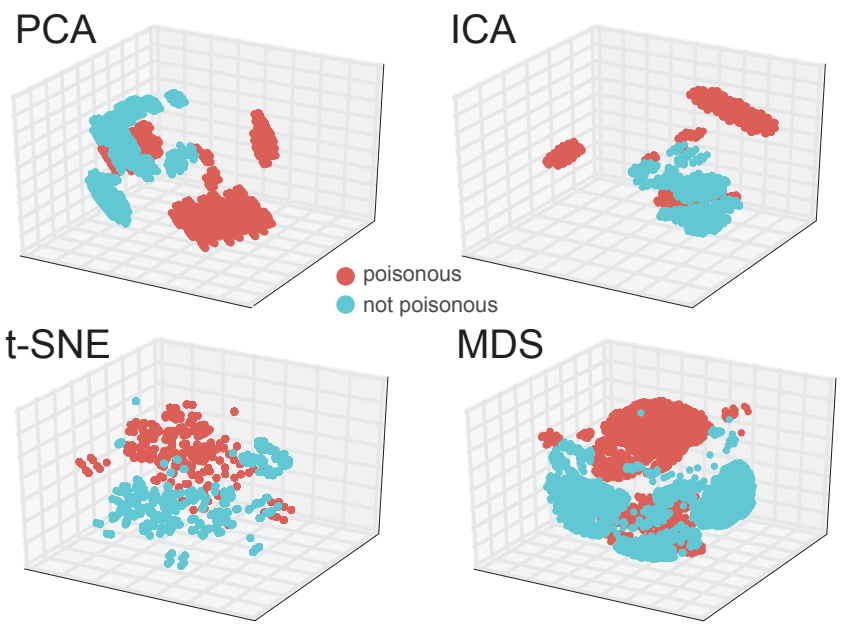


Figure 4.6: Three-dimensional embeddings of the mushrooms dataset using several dimensionality reduction techniques.

Each point represents a sample (mushroom). Red dots indicate poisonous mushrooms and blue indicate non-poisonous mushrooms.

MDS) to visualize their effects on clustering (Fig. 4.6). Each technique produces a unique low-dimensional projection of the data, highlighting distinct structural aspects.

To highlight the sub-clustering structure in this dataset, we use the `n_clusters` argument to `plot`:

```
hyp.plot(mushrooms_data,n_clusters = 23)
```
(4.22)

This command relies on k-means clustering to empirically derive cluster labels, and then plot each cluster in a different color (Fig. 4.7).

Example 0: Exploring factors that influence educational outcomes.

Next, we analyzed an education dataset containing, for each of 480 students around the world, performance ratings (high, medium, and low performance), demographic descrip-

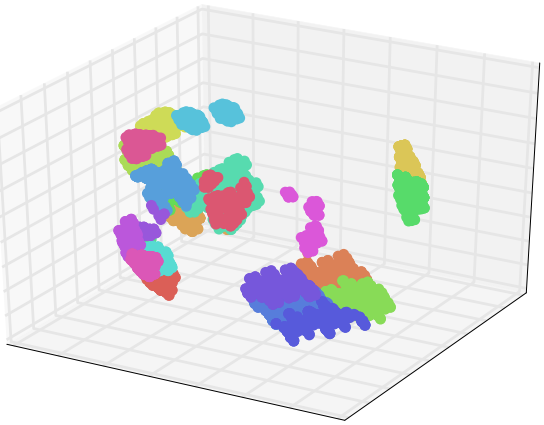


Figure 4.7: Mushrooms dataset, colored by k-means cluster.

	gender	NationalITy	PlaceofBirth	StageID	GradeID	...	Class
0	M	KW	KuwaIT	lowerlevel	G-04	...	M
1	M	KW	KuwaIT	lowerlevel	G-04	...	M
2	M	KW	KuwaIT	lowerlevel	G-04	...	L
3	M	KW	KuwaIT	lowerlevel	G-04	...	L
4	M	KW	KuwaIT	lowerlevel	G-04	...	M

Table 4.3
Example features in education dataset.

The dataset contained categorical and numerical features, as well as student performance labels.

tors (e.g. gender, nationality, place of birth, etc.) as well as classroom behaviors (number of times the student raised their hand, days absent from class, number of times the student visited online resources, etc.) and others (features for five example students are displayed in Tab. 4.3; for full list of features and to download the data, see the Kaggle database). Given a dataframe with the student features, `HyperTools` automatically converts this into a binary data matrix (as described above) for visualization.

In contrast to the mushroom dataset, where the samples formed clear clusters, the distribution of samples in this dataset appear to form a single contiguous mass. Further, coloring each sample (student) by their performance rating reveals a striking correspondence between the student's attributes and performance ratings (Fig. 4.8a). For example,

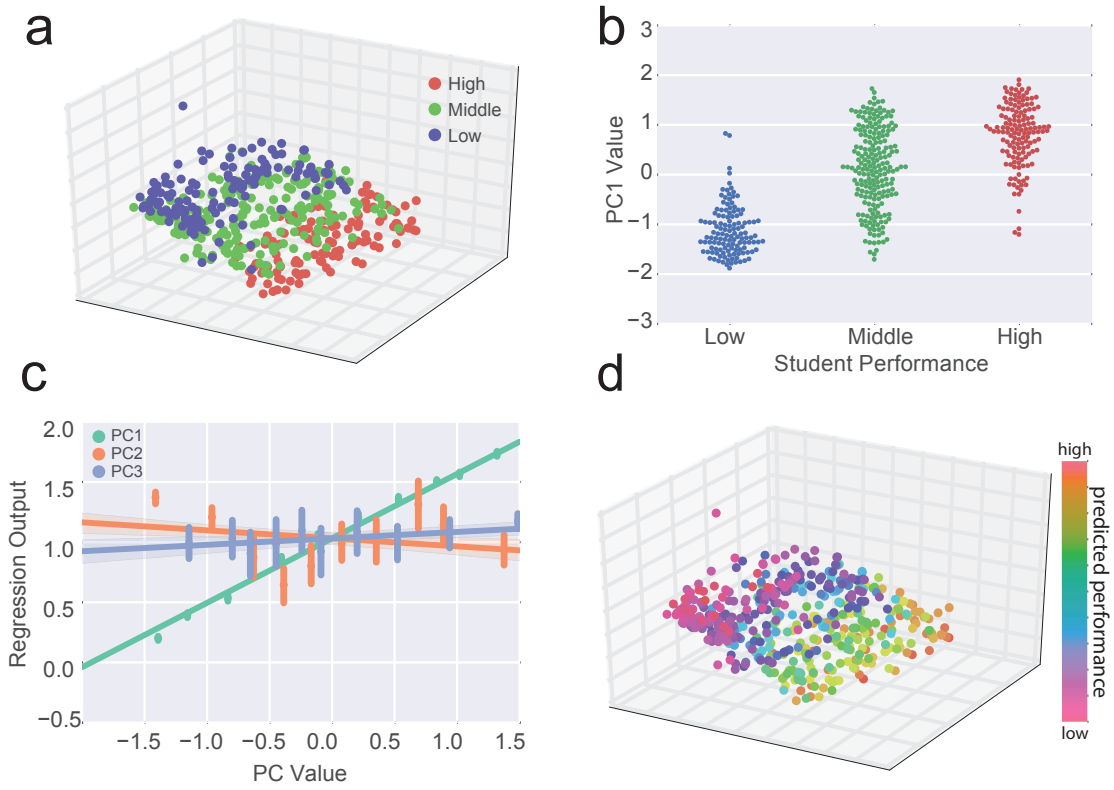


Figure 4.8: Relationship between student attributes and performance.

a. Each point represents the feature vector associated with a student, and the points are colored by student performance (red: high, green: middle, blue: low). **b.** Swarm plot of the first principal component split by student performance (coloring same as above). **c.** Predicted student performance from linear regression of each PC on student performance. **d.** Same as (a), but points are colored by linear regression predictions of student performance by PC1 value (red to violet gradient represents high to low predicted performance).

as the student attributes vary along the first principal component, the student performance ratings appear to transition smoothly from low, to medium, to high (Fig. 4.8b). To highlight this pattern, we fit a linear regression model whose output variable was student performance and the input variables were the first three principal components (Fig. 4.8c). In this way, the regression model's outputs provide a continuous estimate of student performance, whereas the original data contained only discrete (categorized) estimates. In Figure 4.8d, each dot from Panel a has been re-colored according to the regression model's performance predictions, resulting in a smooth gradient from low to high performance.

Example 0: Exploring linguistic data from presidential nominees' Twitter posts.

Whereas the above examples illustrate how simple numerical and categorical features are processed by `HyperTools` to reveal geometric patterns in the data, we can use a similar approach to extract and visualize more complex features. For example, topic models (Blei . 2003) may be used to derive a vector representation of each document in a corpus according to its linguistic properties. Specifically, topic models identify "themes" that are reflected in varying amounts by different documents in the corpus, where each theme (*topic*) is defined formally as a distribution over words in the vocabulary. In other words, a neuroscience-themed topic might heavily weight words like *neuron* and *brain*, whereas a sports-themed topic might heavily weight words like *running* and *athlete*. (Fitting a topic model to a text corpus reveals what the specific topics are and how much each document reflects each topic.) Once we have derived topic vectors for each document in the corpus, we can use `HyperTools` to visualize the full corpus to potentially gain insights into its geometric structure.

As an example of this approach, we next turn to an analysis of Twitter data ("tweets") from the Twitter accounts of Hillary Clinton (@HillaryClinton) and Donald Trump (@re-

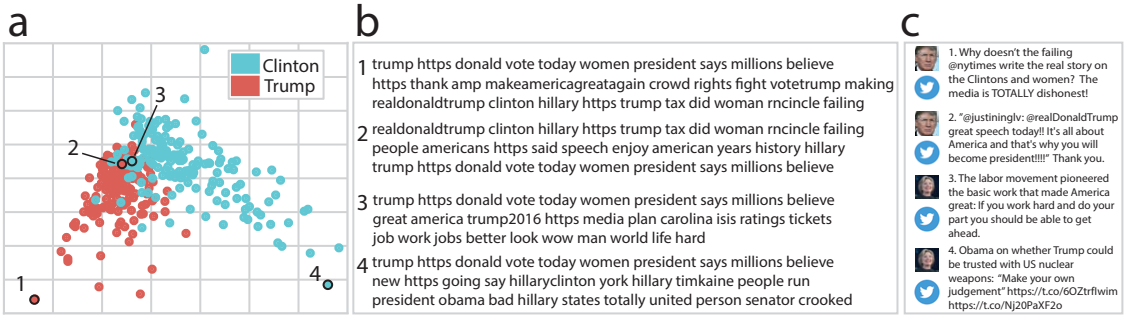


Figure 4.9: Topic models of political Twitter data.

a. Two-dimensional representation of Clinton's (blue) and Trump's (red) day-by-day tweet content. **b.** Top ten words from each of the top three topics on selected days. **c.** Representative tweets from the selected days.

alDonaldTrump) over the course of their 2016 political campaigns. The dataset, sourced from FiveThirtyEight, contains 6,444 tweets sent from the candidates' primary Twitter accounts between April 17, 2016 and September 26, 2016.

We began our analysis by fitting a 20-topic topic model to the entire collection of tweets from both candidates, yielding a single topic vector for each tweet. Separately for each candidate, we next computed daily average topic vectors over the six month interval covered by the dataset, and we used `HyperTools` to visualize the resulting day-by-day Twitter topics.

Plotting the candidates' tweet content in two dimensions reveals that Clinton's and Trump's tweets were primarily about different topics, resulting in a V-like topic cloud (Fig. 4.9a). We leveraged this structure revealed by `HyperTools` to select several days of interest to examine further. Specifically, we examined (1) a day of Trump tweets whose topic coordinates were especially Trump-like (i.e. at the end of the Trump side of the V), (2) a day of Trump tweets whose topic coordinates fell at the intersection of the V, (3) a day of Clinton tweets whose topic coordinates fell at the intersection of the V, and (4) a day of Clinton tweets that fell at the end of the Clinton side of the V. For example, we wondered whether the candidates' tweets that fell at the extreme ends of the V might be especially repre-

sentative of each candidates' unique features, whereas tweets that fell at the intersection of the V might express points of similarity between the candidates. Figure 4.9b displays the top 10 words from each of the top three topics for each of these days of interest, and Figure 4.9c provides representative tweets from each day. Strikingly (perhaps), the most Trump-like tweets appear to disparage Clinton, the most Clinton-like tweets appear to disparage Trump, and the overlapping tweets appear to praise America's greatness.

Example 0: Cyclical increases in global temperatures over time.

In addition to generating static point cloud plots, `HyperTools` may be used to generate trajectory plots to illustrate dynamic patterns in the data. To highlight this feature, we used a global temperatures dataset which we acquired from Berkeley Earth. The Berkeley Earth averaging method takes temperature observations from a large array of weather monitoring stations throughout the world and produces a time-varying estimate of the underlying global temperature field across all of the Earth's land areas. This temperature field may then be sampled to obtain location-specific temperature estimates.

To visualize how the global temperature field changes over time, we acquired monthly average temperature estimates for 20 cities throughout the world (Fig. 4.10c) over the 138 year interval from 1875–2013. We used `HyperTools` to plot the resulting temperature trajectory (Fig. 4.10a,b). To visualize systematic changes over time, we plotted the month-by-month trajectory for each year in a different color using the `group` keyword argument to `plot`:

$$\texttt{hyp.plot(data, group = years, palette = 'RdBu\_r')} \tag{4.23}$$

Two general trends were revealed by plotting the temperature data in this way. First, the month-by-month temperatures within a year are cyclical (e.g. reflecting the changing seasons), which appears in the trajectory as a "figure 8" (this trend is most visible in

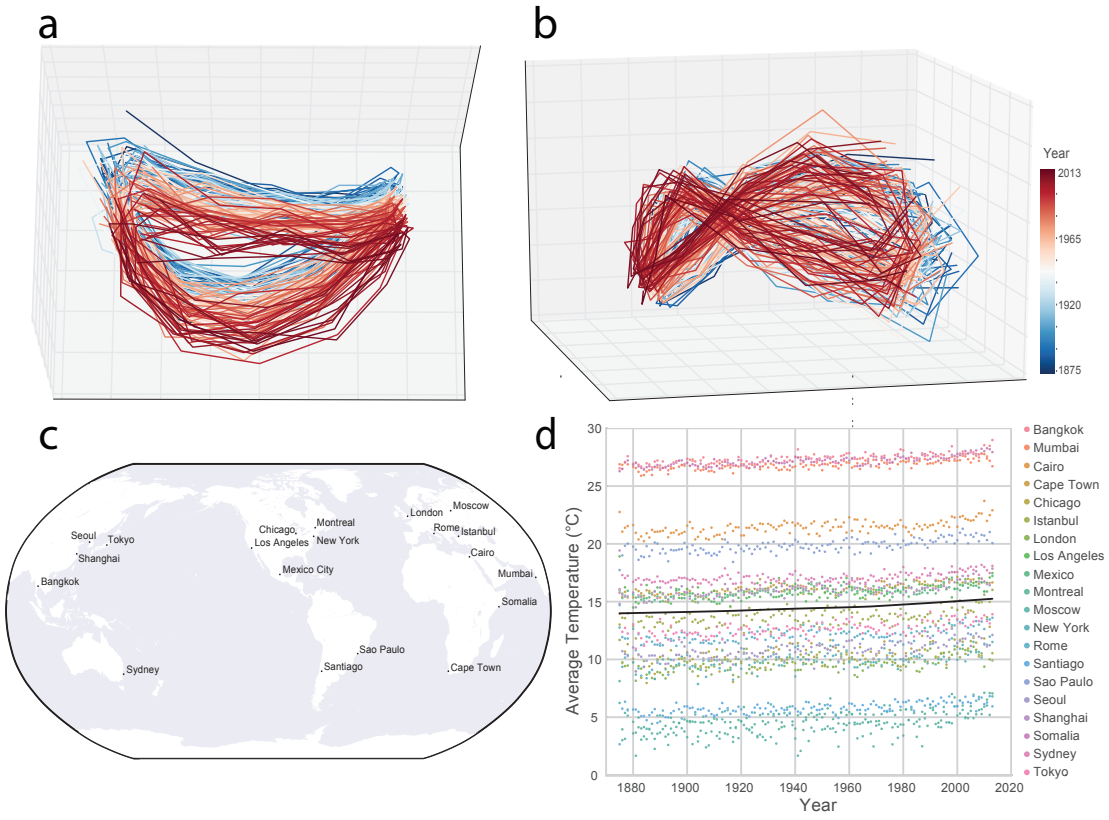


Figure 4.10: Global temperatures from 1875–2013.

a. and **b.** The global temperatures dataset plotted using PCA dimensionality reduction in two views. The line colors change over time (from the earliest time point in blue to the most recent time point in red). The view on the left shows the temporal progression in one of the dimensions while the view on right highlights the cyclical nature of the dataset. **c.** Locations of the 20 cities in the dataset. **d.** Yearly mean temperatures colored by location and black LOWESS line fit to the full dataset.

Fig. 4.10b). Second, there has been a systematic shift in global temperatures over the 138 year period we examined. This appears as a systematic shift in the position of the trajectory over time (Fig. 4.10a), and can also be seen by directly plotting the temperatures over time (Fig. 4.10d).

Example 0: Visualizing the correspondence between neural trajectories and a movie stimulus

In addition to providing plotting tools for visualizing complex data, `HyperTools` also provides tools for aligning trajectories from different sources (see *Align*). For example, suppose we have brain recordings from different people who all watched the same movie. The general shapes of different people's brain data trajectories (showing how everyone's brain responses changed over time while watching the movie), as well as the movie trajectory (showing how the movie itself changed over time), might all share similar properties (e.g. reflecting the covariance structure of the movie and how people responded to it). However, different people's brains may have reflected those similar responses differently, and the dimensions of "brain space" and "movie space" are not directly comparable. As described in *Materials and Methods*, the `HyperTools` toolbox provides an easy-to-use interface for aligning datasets. In this section example we demonstrate some uses of the `align` function using a previously published fMRI dataset (Haxby . 2011), available for download here. The dataset comprises voxel responses from ventral temporal cortex, from each of 11 people, as they watched the feature-length film *Raiders of the Lost Ark*. The data were processed and hyperaligned as described in the original manuscript (Haxby . 2011).

Figure 4.11a displays the trajectory plots for the averaged hyperaligned brain responses from two groups of participants in the original experiment (six in group 1, the remaining five in group 2). The trajectories appear similar in their overall shape (indicating that the two groups of participants had roughly similar brain responses to the movie), but

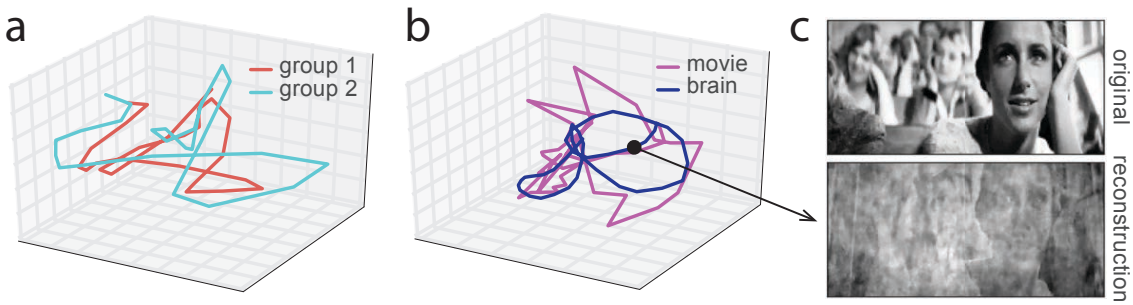


Figure 4.11: Brain/movie trajectories during movie viewing.

a. Group-averaged trajectory of brain activity from ventral visual cortex split into two randomly-selected groups of subjects (group 1: $n = 6$, group 2: $n = 5$) watching the same movie. **b.** Group-averaged trajectory of brain activity from ventral visual cortex and trajectory of movie (pixel intensities over time) hyperaligned to a common space. **c.** Movie frame reconstructed from ventral visual brain activity that is aligned to movie space.

the alignment is imperfect (indicating that understanding individual differences between people's responses might be an interesting future direction to explore).

We next demonstrate how `HyperTools` may be used to visualize the correspondence between datasets with different coordinate systems– specifically, time-varying brain responses to the movie and the time-varying pixel intensities of the movie frames. To align these spaces, we first preprocessed the movie frames to convert the movie into the $S \times F$ matrix format required by `HyperTools` (here S is the number of movie frames and F is the number of pixels per frame). We downsampled the movie frames from 540×960 RGB pixels at 30 FPS to 108×192 grayscale pixels at 1 FPS. We then re-shaped each downsampled frame into a 20,736-dimensional vector.

We next averaged the (hyperaligned) brain responses from the 11 experimental participants to obtain a single brain response matrix. We used piecewise cubic interpolation (Fritsch Carlson 1980) to re-sample this averaged brain response matrix from the original data acquisition rate (one image acquired every 2.5 s) to the downsampled movie frame rate (one image per second). We used the `reduce` function to project both the movie and brain data onto 6,641 dimensions (i.e. the number of voxels in the original brain data) and shifted

the time labels of the brain matrix backwards by 5 s to account for the hemodynamic response. We then used the `procrustes` function to align the brain and movie data:

```
brain_aligned_to_movie = hyp.tools.procrustes(movie_data,brain_data)
```
(4.24)

The resulting aligned brain data matrix may then be plotted in the same space as the movie data matrix (Fig. 4.11b). This visualization can provide insights into the similarities and differences between the geometric structure of the original movie and the structure of the brain responses to the movie.

In addition to facilitating visual comparisons of the geometries of the movie and brain data, the aligned data may also be compared in the "native" data space. For example, each coordinate of "movie space" corresponds to an image, which may be displayed and examined. Aligning the brain data to this movie space (using the `procrustes` function) means that each brain pattern now corresponds to a coordinate in movie space, and therefore the corresponding image may also be displayed and examined (Fig. 4.11c). This provides a means of viewing the original movie through the "lens" of the brain responses to that movie. This general approach could also be carried out in a cross-validated way (i.e. using one portion of the data to compute the Procrustean transformation from brain space to movie space, and then applying that transformation to the held-out brain data). We plan to explore this form of alignment-based decoding in future work.

4.4 Discussion

Visualizing high-dimensional data via low-dimensional embeddings provides an intuitive means of exploring the geometric and statistical properties of complex datasets. This can help to guide analysis decisions and facilitate hypothesis generation and testing. Returning briefly to the example of Anscombe's quartet we discussed in the *Introduction*

(Fig. 4.1), striking differences between datasets with very different geometries may be overlooked when solely considering their summary statistics, and this principle can be extended to high-dimensional data as well. Our `HyperTools` toolbox aims to assist in high-dimensional data visualization by providing a simple (yet powerful) set of plotting functions and data manipulation tools.

We have provided brief examples of how our toolbox may be used to examine data from a wide array of domains: geometry (Example 1), biological data (Example 2), educational and sociological data (Example 3), political and linguistic data (Example 4), and neuroscientific data (Example 5). We chose these particular examples to showcase a broad sampling of the types of visualizations and analyses our toolbox supports, but they are not intended to indicate that our toolbox may be used in only these ways or in these domains.

We hope that `HyperTools` will prove useful in analyzing and visualizing complex data from a wide array of domains. We have released the toolbox under an open-source license to facilitate transparency and widespread adoption. We also hope that users will contribute to the toolbox by providing feedback and suggestions, and by sharing their own extensions and applications with the community.

Conclusion

The way we deploy our attention is constantly, dynamically influencing our experience of the world around us and the aspects of our experience that we store into memory. In this dissertation, I consider visual attention and cognition in two respects: how visual attention interacts with memory, and how we can use our knowledge about visual cognition to inform the ways we visualize, interpret, and communicate scientific findings.

Aim 1 Summary

In Chapters 1 and 2, I explored interactions between the high-level cognitive processes of visual attention and memory. Specifically, I investigated the way feature-based and location-based attention influence memory. I found that feature-based attention suppressed memory for items with unattended features, whereas location-based attention boosted memory for items in attended locations. Notably, these effects were additive but dissociable, operating on independent timescales (with spatial attention operating on a shorter timescale than feature-based). While existing work supports that location-based attention operates at a faster timescale than feature-based attention (Soto Blanco 2004), that location-based attention enhances the processing of attended stimuli whereas feature-based attention suppresses the processing of unattended stimuli and feature-based attention, and that people better remember attended stimuli (Paller Wagner 2002; Chun Turk-Browne 2007; Aly Turk-Browne 2017; Wittig . 2018), prior work has focused largely on elucidating the neural basis of these interactions. My work extends these prior studies by elucidating the specific and separable behavioral impacts of feature-based attention (inhibition with a slow onset) and location-based attention (enhancement with a fast onset) on subsequent memory.

Future Directions

In my study, I found that the effects of feature-based and location-based attention persisted for 2 minutes (through the memory phases of both experiment). As such, one future direction will be to further test the longevity of these effects at longer time intervals. Another important area for future study will be to investigate how the flow of information between different brain structures is modulated by volitional attention (e.g. from sensory regions, such as V1 and A1, to medial temporal lobe structures implicated in encoding experiences into memory. Previous research suggests that attention serves to modulate the *gain* of specific neural circuits (Treue Trujillo 1999; Chance . 2002; Eldar . 2013; Salinas Thier 2000), thus facilitating or inhibiting the flow of specific neural representations (Vartanian . 2007; LaRocque . 2014). Specifcally, feature-based attention may be supported by changes in connectivity with the thalamus (Schneider 2011), whereas location-based attention may be supported by changes in connectivity with primary visual cortex (Noudoost . 2010). Feature-based and location-based attention being mediated by different brain structures may explain why these two types of attention affect memory differently. One way to test this will be to measure neural activity patterns as people modulate their focus of attention (e.g., using functional magnetic resonance imaging or electroencephalography), then using neural decoding approaches (e.g., Haxby . 2001; Manning . 2018) to follow how neural representations of attended and unattended stimuli ascend from primary sensory regions, to higher order sensory regions, and, finally, to memory areas. We would predict that, if the effects of attention on memory are mediated by changes in network dynamics, then transmission of representations of attended stimuli might be facilitated, with variability in neural changes tracking with behavioral measures of memorability.

Aim 2 Summary

Next, In Aim 2 (Chapters 3 and 4), I developed tools to facilitate the intuitive inter-

pretation of complex and high-dimensional data. I started by analyzing the capabilities of modern speech-to-text algorithms to transcribe verbal data from psychology experiments. I found that speech-to-text algorithms transcribe accurately and efficiently, and, critically, that automatic transcriptions preserve pivotal aspects of the data relevant to studies of human memory. Second, I used data reduction and alignment algorithms along with visualization techniques to explore high-dimensional data. This yielded an open-source software package (HyperTools) for manipulating and visualizing high-dimensional data. In the future, the Hypertools framework can be used for a wide variety of applications to cognitive data, affording novel insights into high-dimensional psychological phenomena. For example, one future direction for the project will be to use Hypertools to explore clinically relevant relationships between cognitive and psychiatric data.

Future Directions

Mental illness continues to be one of the major challenges of the 21st century. Fortunately, recent advances in psychology and computer science are ideally suited to the early, data-driven detection of mental illness. Psychology research has revealed that mental illness is often associated with (or preceded by) subtle changes in baseline cognition and perception. Identifying, aggregating, and interpreting these cognitive changes would be intractable for classical psychologists, but it is precisely the type of problem machine learning is designed to solve. Specifically, my recently published software, Hypertools (Heusser et al., 2018), makes it easy to apply novel combinations of powerful machine learning analyses and data visualization methods. Recall that the overarching goal of the toolbox is to provide a computational framework for gaining visual insights into high dimensional data, with a key feature of this framework being the ability to compare patterns in data collected across modalities (e.g. behavioral, neural, and psychiatric datasets).

As such, this framework is ideally suited to help identify cognitive warning signs of mental illness by relating high-dimensional cognitive and psychiatric data. Specifi-

cally, we can use HyperTools to reduce and map rich behavioral data (from a battery of perception, memory, and cognitive control tasks; Table 1) onto psychiatric space (spanning symptoms associated with eight classical psychiatric disorders; Table 2). By reducing high-dimensional cognitive data and high-dimensional psychiatric data in an initial sample of participants, we can estimate the optimal mapping of datapoints in cognitive space to points in psychiatric space. If we are able to estimate this mapping well, and an optimal mapping indeed exists (as we hypothesize it does), then we should be able to apply this mapping to new and held-out cognitive data in a way that reliable predicts psychiatric data. That is, we can quantify a person's cognitive traits and use these to predict their psychiatric traits, by reducing their cognitive data and mapping it into a three dimensional psychiatric space.

This approach to mental illness detection has numerous benefits to offer, above and beyond current standard approaches. First, since it is based on quantitative analysis rather than subjective assessment, cognitive-psychiatric mappings can help disambiguate cases where physicians' subjective assessments disagree. Also, since machine learning algorithms can capitalize on patterns in data that are too subtle or complex for human interpreters to recognize, cognitive-psychiatric mappings have the potential to detect the onset of mental illness earlier, based on initial corresponding changes in cognitive data that may be too subtle or complex to be recognized by human interpreters.

Concluding remarks

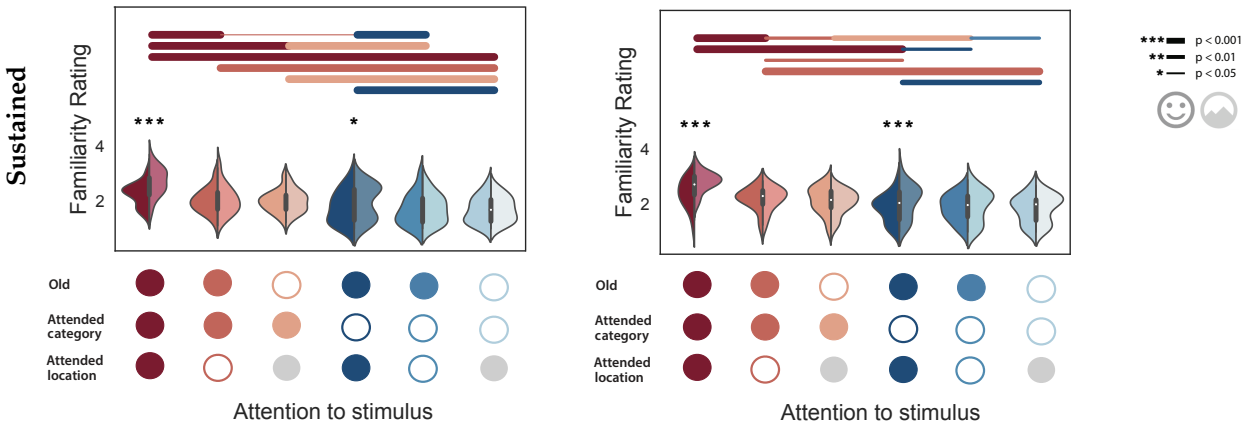
My work has revealed insights into the interactive processes of visual attention and recognition memory. Further, I have worked to use what we know about visual cognition to facilitate easier interpretation, understanding, and communication of scientific data. I hope that the findings and interpretive frameworks outlined in this dissertation will continue to meaningfully contribute to to wide reaching applications, including mental health initiatives, in years to come.

Appendix A

Supporting Materials for *Feature-based and location-based volitional covert attention affect memory at different timescales*

(a) Initial cohort

(b) Replication cohort



(c) Initial cohort

(d) Replication cohort

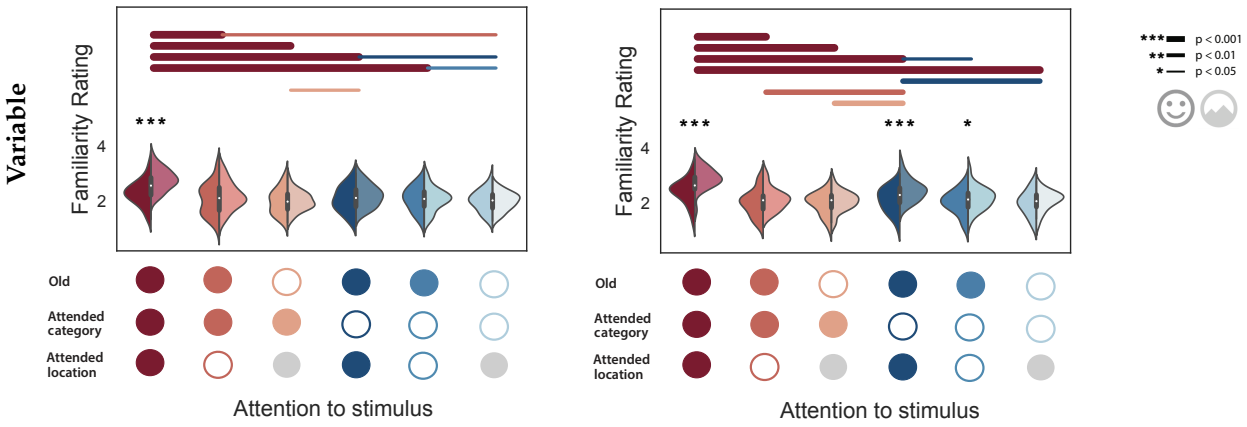
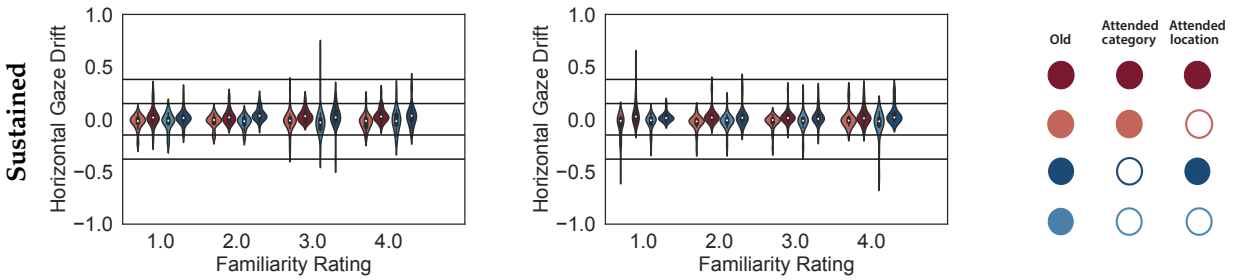


Figure S1: **Familiarity ratings for attended, unattended, and novel stimuli.** This figure follows the same formatting as Figure 2 in the main text, but the panels in this figure break down those results by participant cohort. Each split violin plot displays the distribution of within-participant average familiarity ratings given to faces (left, darker colors) and scenes (right, lighter colors) during the memory phases of each experiment. As shown in the legend (bottom), the colors indicate whether each image had been viewed during the presentation phase (`old`) or not; whether the given images matched the `attended category`; and/or whether the given images matched the `attended location`. The colored lines above each set of violin plots denote statistical differences (**pos**itive or **neg**ative differences in mean, collapsing over image category, assessed via two-tailed *t*-tests) between the distributions centered on the endpoints of each line. The line thicknesses denote *p*-values as indicated in the legend. Asterisks denote differences between the face versus scene distributions (assessed via two-tailed *t*-tests).

(a) Initial cohort

(b) Replication cohort



(c) Initial cohort

(d) Replication cohort

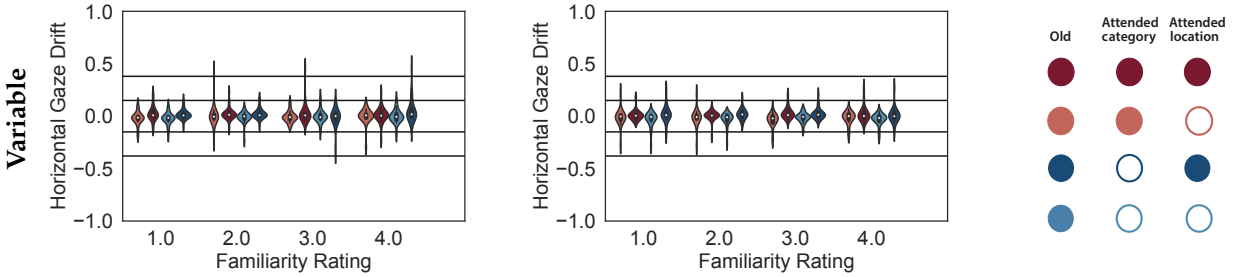


Figure S2: **Horizontal coordinates of visual fixations.** This figure follows the same formatting as Figure 3 in the main text, but the panels in this figure break down those results by participant cohort. Each violin plot displays a distribution of participant-wise average horizontal gaze positions relative to a presented image. The coordinates have been normalized such that a value of 1.0 denotes the furthest onscreen coordinate from the central fixation point *towards* the direction of the given image, and -1.0 denotes the furthest onscreen coordinate from the central fixation point *away from* the direction of the given image. The gray bars in each panel mark the boundaries of the presented images. The violin plots are broken down by participants' familiarity ratings during the memory phases of each experiment, and by each image's relation to the attention cue while that image appeared onscreen.

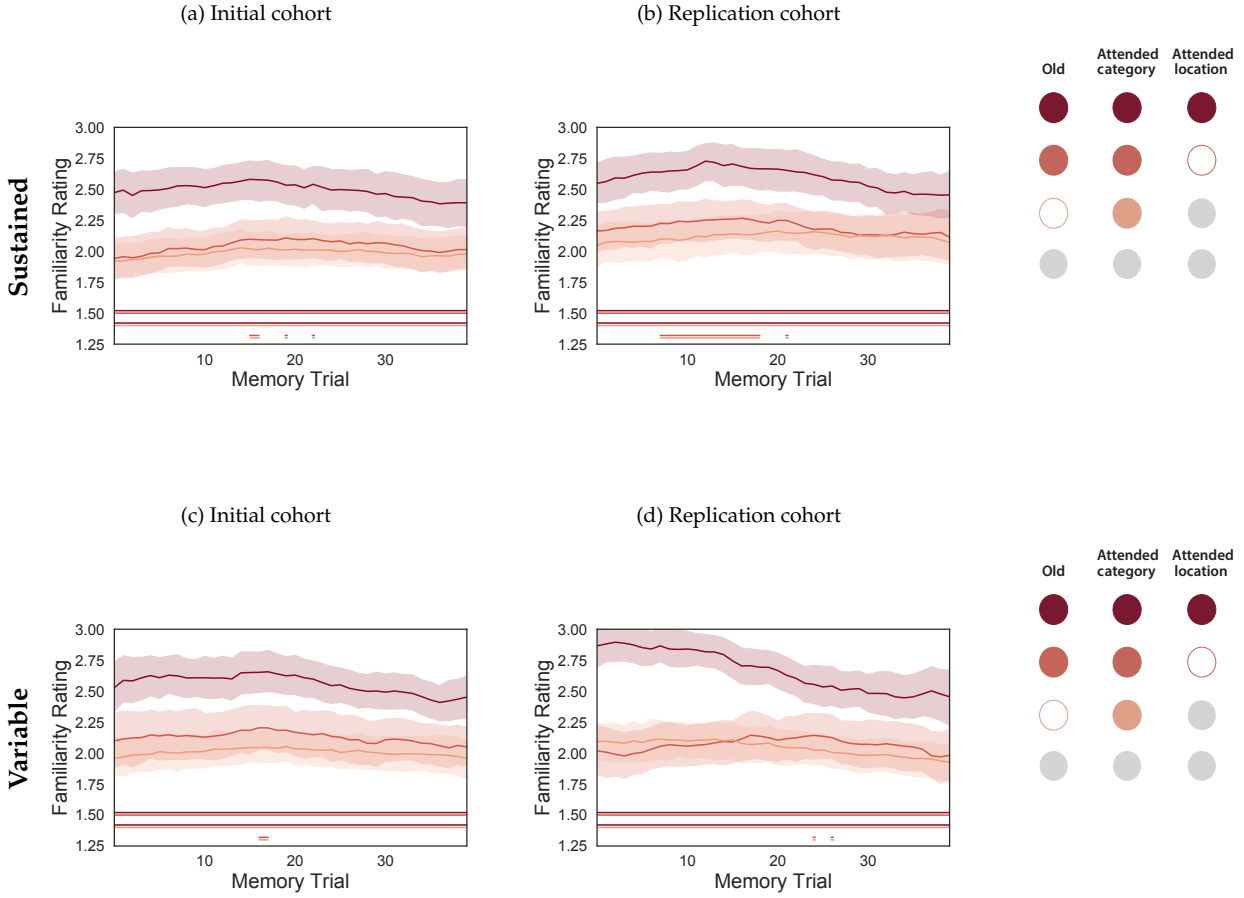


Figure S3: **Familiarity ratings over time for images that matched cued image category.** This figure follows the same formatting as Figure 4 (top panels) in the main text, but the panels in this figure break down those results by participant cohort. Each curve reflects the average familiarity ratings for attended, unattended, and novel images (denoted in the legends on the right) within a succession of overlapping 20-image sliding windows. Error ribbons denote 95% confidence intervals, computed across participants. The paired horizontal lines at the bottom of each panel denote timepoints when the given pair of curves was statistically distinguishable (i.e., the topmost line color was statistically greater than the bottommost line color at $\alpha = 0.05$, via a paired two-tailed t-test.)

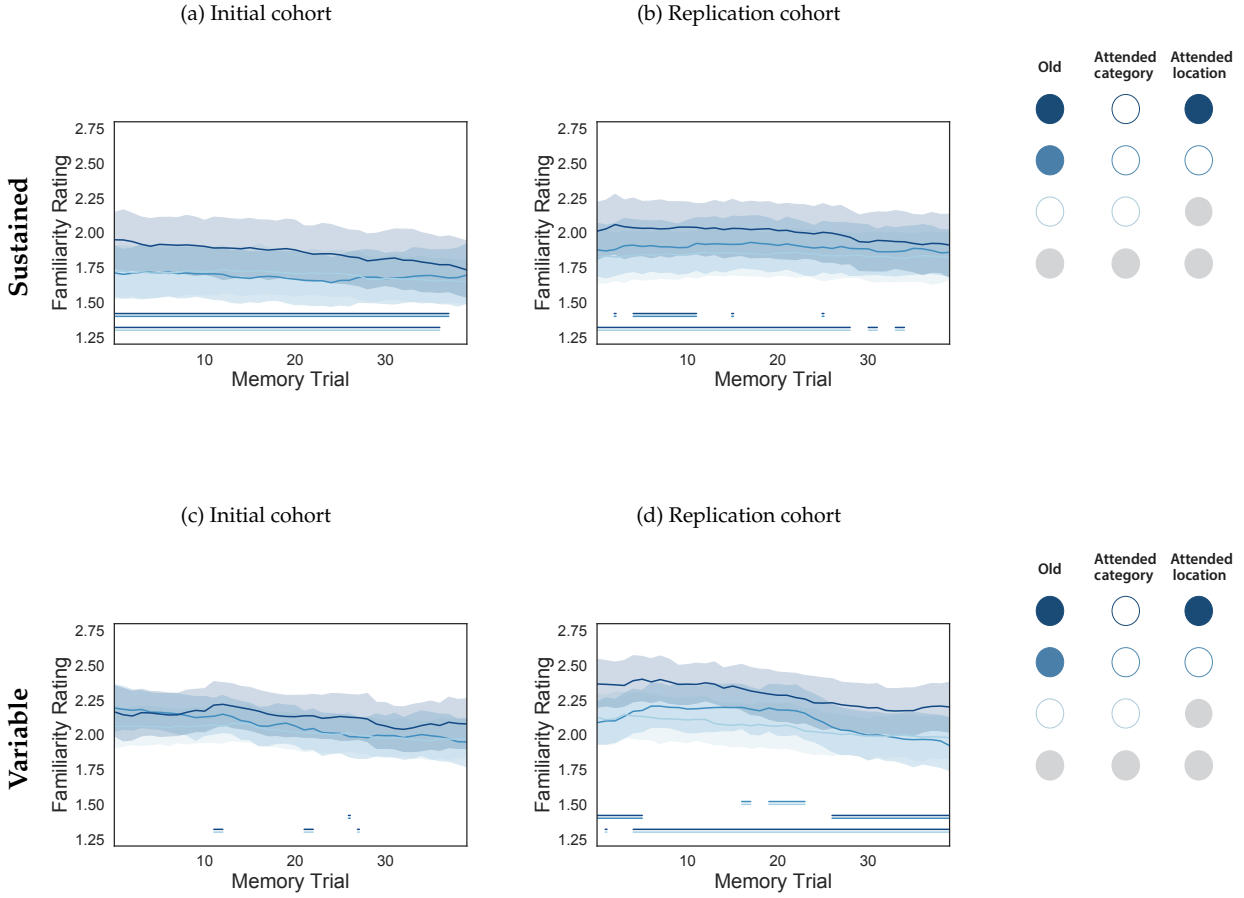
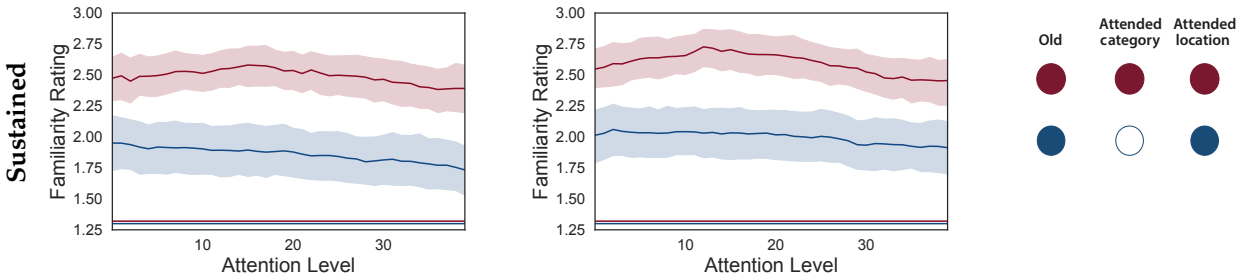


Figure S4: **Familiarity ratings over time for images that matched uncued image category.** This figure follows the same formatting as Figure 4 (bottom panels) in the main text, but the panels in this figure break down those results by participant cohort. Each curve reflects the average familiarity ratings for attended, unattended, and novel images (denoted in the legends on the right) within a succession of overlapping 20-image sliding windows. Error ribbons denote 95% confidence intervals, computed across participants. The paired horizontal lines at the bottom of each panel denote timepoints when the given pair of curves was statistically distinguishable (i.e., the topmost line color was statistically greater than the bottommost line color at $\alpha = 0.05$, via a paired two-tailed t-test.)

(a) Initial cohort

(b) Replication cohort



(c) Initial cohort

(d) Replication cohort

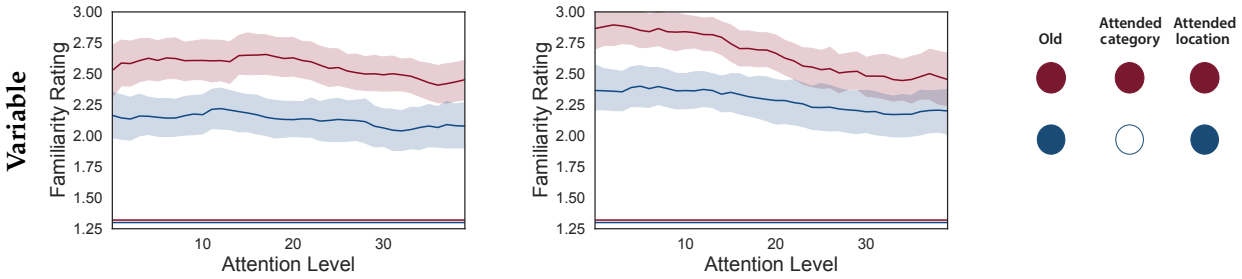
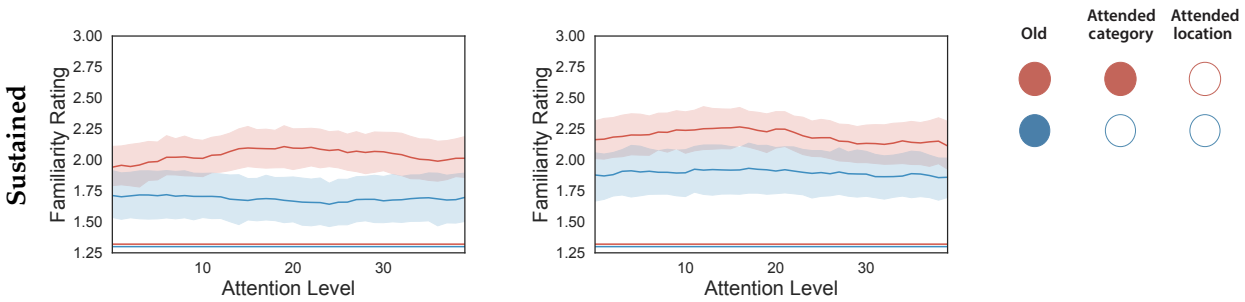


Figure S5: **Familiarity ratings over time for images that matched the cued image location.** This figure follows the same formatting as Figure 5 (top panels) in the main text, but the panels in this figure break down those results by participant cohort.

(a) Initial cohort

(b) Replication cohort



(c) Initial cohort

(d) Replication cohort

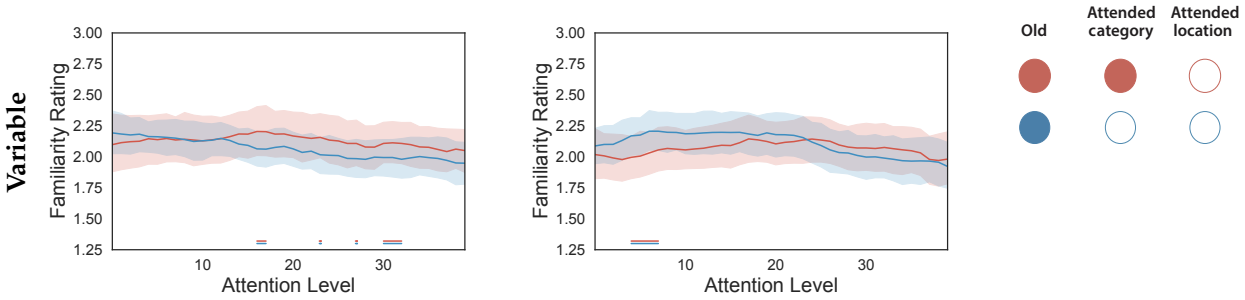
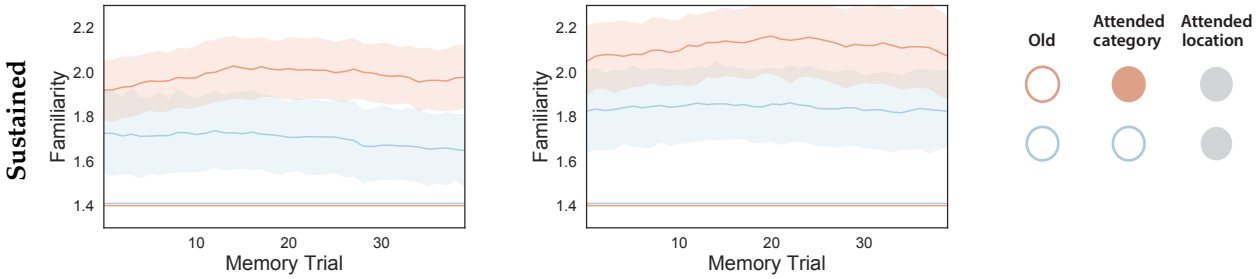


Figure S6: **Familiarity ratings over time for images that matched uncued image location.** This figure follows the same formatting as Figure 5 (bottom panels) in the main text, but the panels in this figure break down those results by participant cohort.

(a) Initial cohort

(b) Replication cohort



(c) Initial cohort

(d) Replication cohort

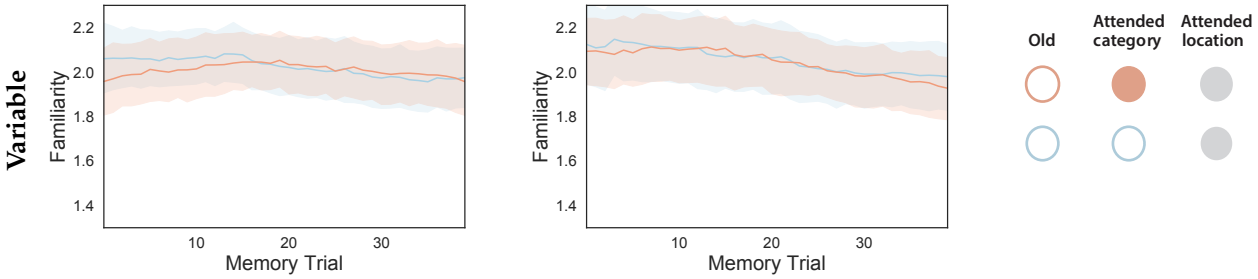
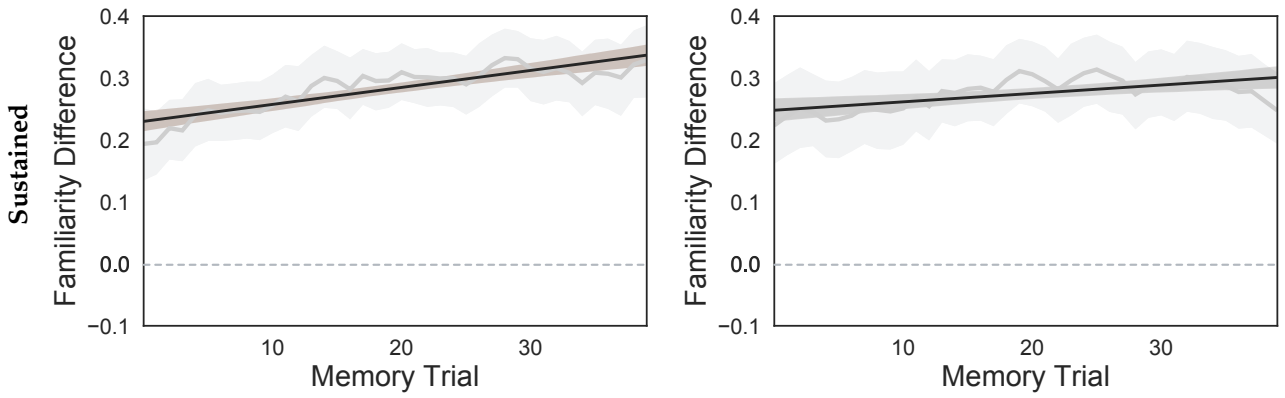


Figure S7: **Familiarity ratings over time for attended-category and unattended-category novel images.** This figure follows the same formatting as Figure 6 (top panels) in the main text, but the panels in this figure break down those results by participant cohort.

(a) Initial cohort

(b) Replication cohort



(c) Initial cohort

(d) Replication cohort

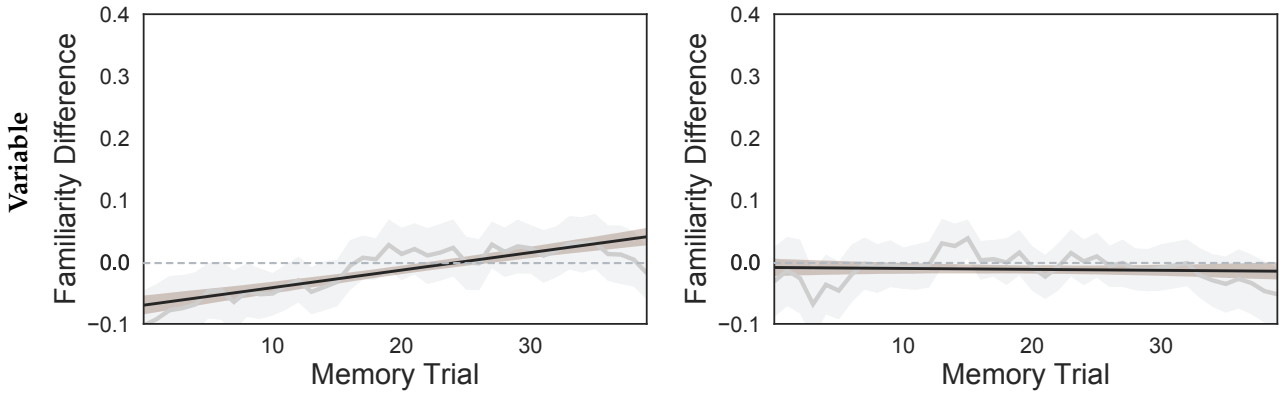


Figure S8: **Familiarity ratings over time for attended-category and unattended-category novel images.** This figure follows the same formatting as Figure 6 (bottom panels) in the main text, but the panels in this figure break down those results by participant cohort.

Bibliography

Ahern, S. Beatty, J. 1981. Physiological evidence that demand for processing capacity varies with intelligence Physiological evidence that demand for processing capacity varies with intelligence. Intelligence and Learning Intelligence and learning (121–128).

Aly, M. Turk-Browne, NB. 2017. How hippocampal memory shapes, and is shaped by, attention How hippocampal memory shapes, and is shaped by, attention. The Hippocampus From Cells to Systems The hippocampus from cells to systems (369–403). Springer.

Angelakis, E., Stathopoulou, S., Frymiare, JL., Green, DL., Lubar, JF. Kounios, J. 2007. EEG neurofeedback: a brief overview and an example of peak alpha frequency training for cognitive enhancement in the elderly EEG neurofeedback: a brief overview and an example of peak alpha frequency training for cognitive enhancement in the elderly. The Clinical Neuropsychologist211110–129.

Anscombe, FJ. 1973. Graphs in statistical analysis Graphs in statistical analysis. American Statistitian27117–21.

Bamberg, P., lu Chow, Y., Gillick, L., Roth, R. Sturtevant, D. 1990. The Dragon continuous speech recognition system: a real-time implementation The Dragon continuous speech recognition system: a real-time implementation. Proceedings of the

DARPA Speech and Natural Language Workshop Proceedings of the DARPA speech and natural language workshop (78–81).

Beukema, S., Jennings, BJ., Olson, JA. Kingdom, FAA. 2019. The pupillary response to the unknown: novelty versus familiarity The pupillary response to the unknown: novelty versus familiarity. i-Perception1051–12.

Blei, DM., Ng, AY. Jordan, MI. 2003. Latent dirichlet allocation Latent dirichlet allocation. Journal of Machine Learning Research3993–1022.

Buhrmester, M., Kwang, T. Gosling, SD. 2011. Amazon's Mechanical Turk: a new source of inexpensive, yet high-quality, data? Amazon's Mechanical Turk: a new source of inexpensive, yet high-quality, data? Perspectives on Psychological Science613–5.

Carlini, N. Wagner, D. 2018. Audio adversarial examples: targeted attacks on speech-to-text Audio adversarial examples: targeted attacks on speech-to-text. arXiv1801.01944.

Chance, FS., Abbott, LF. Reyes, AD. 2002. Gain modulation from background synaptic input Gain modulation from background synaptic input. Neuron354773–782.

Chang, LJ., Jolly, E., Cheong, JH., Rapuano, K., Greenstein, N., Chen, PH. Manning, JR. 2018. Endogenous variation in ventromedial prefrontal cortex state dynamics during naturalistic viewing reflects affective experience Endogenous variation in ventromedial prefrontal cortex state dynamics during naturalistic viewing reflects affective experience. bioRxivdoi.org/10.1101/487892.

Chen, PH., Chen, J., Yeshurun, Y., Hasson, U., Haxby, J. Ramadge, PJ. 2015. A Reduced-Dimension fMRI Shared Response Model A Reduced-Dimension fMRI Shared Response Model. C. Cortes and N. D. Lawrence and D. D. Lee and M. Sugiyama and R.

Garnett (), Advances in Neural Information Processing Systems 28 Advances in Neural Information Processing Systems 28 (460–468). Curran Associates, Inc.

Chun, MM. Turk-Browne, NB. 2007. Interactions between attention and memory Interactions between attention and memory. Current Opinion in Neurobiology172177–184.

Cohen, MS. 2001. Real-time functional magnetic resonance imaging Real-time functional magnetic resonance imaging. Methods25201–220.

Col, J. 2017. Enchanted Learning. Enchanted learning. http://www.enchantedlearning.com

Colizoli, O., de Gee, JW., Urai, AE. Donner, TH. 2018. Task-evoked pupil responses reflect internal belief states Task-evoked pupil responses reflect internal belief states. Scientific Reports8137021–13.

Comon, P., Jutten, C. Herault, J. 1991. Blind separation of sources, part II: problems statement Blind separation of sources, part II: problems statement. Signal Processing24111–20.

Cornsweet, TN. 1962. The staircase-method in psychophysics The staircase-method in psychophysics. American Journal of Psychology753485–491.

Cox, RW. Jesmanowicz, A. 1999. Real-time 3D image registration for functional MRI Real-time 3D image registration for functional MRI. Magnetic Resonance in Medicine421014–1018.

Cox, RW., Jesmanowicz, A. Hyde, JS. 1995. Real-time functional magnetic resonance imaging Real-time functional magnetic resonance imaging. Magnetic Resonance in Medicine33230–236.

Crump, MJC., McDonnell, JV. Gureckis, TM. 2013. Evaluating Amazon's Mechanical Turk as a tool for experimental behavioral research Evaluating Amazon's Mechanical Turk as a tool for experimental behavioral research. PLoS One83e57410.

de Leeuw, JR. 2015. jsPsych: A JavaScript library for creating behavioral experiments in a web browser jsPsych: A JavaScript library for creating behavioral experiments in a web browser. Behavior Research Methods4711–12.

deBettencourt, MT., Cohen, JD., Lee, RF., Norman, KA. Turk-Browne, NB. 2015. Closed-loop training of attention with real-time brain imaging Closed-loop training of attention with real-time brain imaging. Nature Neuroscience183470–475.

deCharms, RC. 2008. Applications of real-time fMRI Applications of real-time fMRI. Nature Reviews Neuroscience99720–729.

Deese, J. 1959. On the prediction of occurrence of particular verbal intrusions in immediate recall On the prediction of occurrence of particular verbal intrusions in immediate recall. Journal of Experimental Psychology: General5817–22.

Dijksterhuis, A. Aarts, H. 2010. Goals, attention, and (un)consciousness Goals, attention, and (un)consciousness. Annual Review of Psychology61467–490.

Einhäuser, W. 2017. The pupil as marker of cognitive processes The pupil as marker of cognitive processes. Computational and Cognitive Neuroscience of Vision Computational and cognitive neuroscience of vision (141–169). Springer.

Eldar, E., Cohen, JD. Niv, Y. 2013. The effects of neural gain on attention and learning The effects of neural gain on attention and learning. Nature Neuroscience1681146.

Engbert, R. Kliegl, R. 2003. Microsaccades uncover the orientation of covert attention Microsaccades uncover the orientation of covert attention. Vision Research4391035–1045.

Faber, NJ. 2017. Neuromodulation of pupil diameter and temporal perception Neuromodulation of pupil diameter and temporal perception. The Journal of Neuroscience37112806–2808.

Fiedler, S. Glöckner, A. 2012. The dynamics of decision making in risky choice: an eye-tracking analysis The dynamics of decision making in risky choice: an eye-tracking analysis. Frontiers in Psychology3OCT1–18.

Freeman, J. Simoncelli, EP. 2011. Metamers of the ventral stream Metamers of the ventral stream. Nature Neuroscience141195–1201.

Friendly, M. 2006. A brief history of data visualization A brief history of data visualization. C. Chen, W. Härdle A. Unwin (), Handbook of Computational Statistics: Data Visualization Handbook of computational statistics: Data visualization (III). Heidelberg, GermanySpringer.

Fritsch, FN. Carlson, RE. 1980. Monotone piecewise cubic interpolation Monotone piecewise cubic interpolation. Society for Industrial and Applied Mathematics Journal on Numerical Analysis172238–246.

Gallo, D. 2006. Associative illusions of memory: false memory research in DRM and related tasks Associative illusions of memory: false memory research in DRM and related tasks. New York, NYPsychology Press.

Gardner, RM., Beltramo, JS. Krinsky, R. 1975. Pupillary changes during encoding, storage, and retrieval of information Pupillary changes during encoding, storage, and retrieval of information. Perceptual and Motor Skills413951–955.

Gardner, RM., Mo, SS. Borrego, R. 1974. Inhibition of pupillary orienting reflex by novelty in conjunction with recognition memory Inhibition of pupillary orienting reflex by novelty in conjunction with recognition memory. Bulletin of the Psychonomic Society33237–238.

Godden, DR. Baddeley, AD. 1975. Context-dependent memory in two natural environments: on land and under water Context-dependent memory in two natural environments: on land and under water. British Journal of Psychology66325–331.

Goldinger, SD. Papesh, MH. 2012. Pupil dilation reflects the creation and retrieval of memories Pupil dilation reflects the creation and retrieval of memories. Current Directions in Psychological Science21290–95.

Gureckis, TM., Martin, J., McDonnell, J., Rich, AS., Markant, D., Coenen, A.Chan, P. 2015. psiTurk: An open-source framework for conducting replicable behavioral experiments online psiTurk: An open-source framework for conducting replicable behavioral experiments online. Behavior Research Methods483829–842.

Haj, ME., Janssen, SMJ., Gallouj, K. Lenoble, Q. 2019. Autobiographical memory increases pupil dilation Autobiographical memory increases pupil dilation. Translational Neuroscience10280–287.

Halpern, Y., Hall, KB., Schogol, V., Riley, M., Roark, B., Skobeltsyn, G. Bäuml, M. 2016. Contextual prediction models for speech recognition Contextual prediction models for speech recognition. Interspeech Interspeech (2338–2342).

Hardt, O. Nadel, L. 2009. Cognitive maps and attention Cognitive maps and attention. Progress in Brain Research176181–194.

Hartigan, JA. Wong, MA. 1979. Algorithm AS 136: A k-means clustering algorithm Algorithm AS 136: A k-means clustering algorithm. Journal of the Optical Society of America281100–108.

Haxby, JV., Gobbini, MI., Furey, ML., Ishai, A., Schouten, JL. Pietrini, P. 2001. Distributed and overlapping representations of faces and objects in ventral temporal cortex Distributed and overlapping representations of faces and objects in ventral temporal cortex. Science2932425–2430.

Haxby, JV., Guntupalli, JS., Connolly, AC., Halchenko, YO., Conroy, BR., Gobbini, MI.Ramadge, PJ. 2011. A common, high-dimensional model of the representational space in human ventral temporal cortex A common, high-dimensional model of the representational space in human ventral temporal cortex. Neuron72404–416.

Heaver, B. Hutton, SB. 2011. Keeping an eye on the truth? pupil size changes associated with recognition memory Keeping an eye on the truth? pupil size changes associated with recognition memory. Memory194398–405.

Heusser, AC., Fitzpatrick, PC., Field, CE., Ziman, K. Manning, JR. 2017. Quail: a Python toolbox for analyzing and plotting free recall data Quail: a Python toolbox for analyzing and plotting free recall data. The Journal of Open Source Software10.21105/joss.00424.

Hinton, G. 2012. Neural networks for machine learning Neural networks for machine learning. Courserahttps://www.class-central.com/mooc/398/coursera-neural-networks-for-machine-learning.

Hinton, G., Deng, L., Yu, D., Dahl, GE., Mohamed, AR., Jaitly, N.Kingsbury, B. 2012. Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups. IEEE Signal Processing Magazine29682–97.

Hoffman, JE. Subramaniam, B. 1995. The role of visual attention in saccadic eye movements The role of visual attention in saccadic eye movements. Perception and Psychophysics576787–795.

Huggins-Daines, D., Kumar, M., Chan, A., Black, AW., Ravishankar, M. Rudnicky, AI. 2006. Pocketsphinx: A free, real-time continuous speech recognition system for hand-held devices Pocketsphinx: A free, real-time continuous speech recognition system for hand-held devices. IEEE Conference on Acoustics, Speech, and Signal Processing IEEE conference on acoustics, speech, and signal processing (1, 185–188).

Hunt, AR. Kingstone, A. 2003. Covert and overt voluntary attention: linked or independent? Covert and overt voluntary attention: linked or independent? Cognitive Brain Research181102–105.

Hunt, E., Pellegrino, JW. Yee, PL. 1989. Individual differences in attention Individual differences in attention. Psychology of Learning and Motivation24285–310.

Hunter, JD. 2007. Matplotlib: A 2D graphics environment Matplotlib: A 2D graphics environment. Computing in Science and Engineering9390–95.

Jacoby, LL. 1991. A process dissociation framework: separating automatic from intentional uses of memory A process dissociation framework: separating automatic from intentional uses of memory. Journal of Memory and Language30513–541.

Jacoby, LL., Lindsay, DS. Toth, JP. 1992. Unconscious influences revealed: attention, awareness, and control Unconscious influences revealed: attention, awareness, and control. American Psychologist476802–809.

Jutten, C. Herault, J. 1991. Blind separation of sources, part I: an adaptive algorithm based on neuromimetic architecture Blind separation of sources, part I: an adaptive algorithm based on neuromimetic architecture. Signal Processing2411–10.

Kafkas, A. Montaldi, D. 2015. The pupillary response discriminates between subjective and objective familiarity and novelty The pupillary response discriminates between subjective and objective familiarity and novelty. Psychophysiology52101305–1316.

Kahana, MJ. 1996. Associative retrieval processes in free recall Associative retrieval processes in free recall. Memory and Cognition24103–109.

Kahana, MJ. 2012. Foundations of human memory Foundations of human memory. New York, NYOxford University Press.

Kahana, MJ. 2017. Memory search Memory search. JH. Byrne (), Learning and Memory: A Comprehensive Reference Learning and memory: A comprehensive reference (181–200). Oxford University Press.

Kahneman, D. Beatty, J. 1966. Pupil diameter and load on memory Pupil diameter and load on memory. Science15437561583–1585.

Kahneman, D., Beatty, J. Pollack, I. 1967. Perceptual deficit during a mental task Perceptual deficit during a mental task. Science1573785218–219.

Kang, O. Wheatley, T. 2015. Pupil dilation patterns reflect the contents of consciousness Pupil dilation patterns reflect the contents of consciousness. Consciousness and Cognition35128–135.

Kang, O. Wheatley, T. 2017. Pupil dilation patterns spontaneously synchronize across individuals during shared attention Pupil dilation patterns spontaneously synchronize across individuals during shared attention. Journal of Experimental Psychology: General1464569–576.

Kang, OE., Huffer, KE. Wheatley, TP. 2014. Pupil dilation dynamics track attention to high-level information Pupil dilation dynamics track attention to high-level information. PLoS One98e102463.

Kloosterman, NA., Meindertsma, T., van Loon, AM., Lamme, VAF., Bonneh, YS. Donner, TH. 2015. Pupil size tracks perceptual content and surprise Pupil size tracks perceptual content and surprise. European Journal of Neuroscience4181068–1078.

Korn, CW. Bach, DR. 2016. A solid frame for the window on cognition: modeling event-related pupil responses A solid frame for the window on cognition: modeling event-related pupil responses. Journal of Vision16328.

Kucewicz, MT., Berry, BM., Miller, LR., Khadjevand, F., Ezzyat, Y., Stein, JM.Worrell, GA. 2018. Evidence for verbal memory enhancement with electrical brain stimulation in the lateral temporal cortex Evidence for verbal memory enhancement with electrical brain stimulation in the lateral temporal cortex. Brain1414971–978.

Kurzweil, R., Richter, R., Kurzweil, R. Schneider, ML. 1990. The age of intelligent machines The age of intelligent machines. CambridgeMIT Press.

Laeng, B. Endestad, T. 2012. Bright illusions reduce the eye's pupil Bright illusions reduce the eye's pupil. Proceedings of the National Academy of Sciences, USA10962162–2167.

Laeng, B., Sirois, S. Gredeback, G. 2012. Pupillometry: A window to the preconscious? Pupillometry: A window to the preconscious? Perspectives on Psychological Science7118–27.

LaRocque, JJ., Lewis-Peacock, JA. Postle, BR. 2014. Multiple neural states of representation in short-term memory? It's a matter of attention Multiple neural states of representation in short-term memory? it's a matter of attention. Frontiers in Human Neuroscience85.

Lewis, MB. 2010. Familiarity, target set and false positives in face recognition Familiarity, target set and false positives in face recognition. European Journal of Cognitive Psychology94437–459.

Lincoff, G. National Audubon Society. 1981. The Audubon Society Field Guide to North American Mushrooms The audubon society field guide to north american mushrooms. Knopf. https://books.google.com/books?id=bf8UAQAAIAAJ

Loftus, EF. 1997. Creating false memories Creating false memories. Scientific American277370–75.

Madry, A., Makelov, A., Schmidt, L., Tsipras, D. Vladu, A. 2017. Towards deep learning models resistant to adversarial attacks Towards deep learning models resistant to adversarial attacks. arXiv1706.06083.

Mandler, G. 1980. Recognizing: the judgment of previous occurrence Recognizing: the judgment of previous occurrence. Psychological Review87252–271.

Manning, JR., Norman, KA. Kahana, MJ. 2015. The role of context in episodic memory The role of context in episodic memory. M. Gazzaniga (), The Cognitive Neurosciences The cognitive neurosciences (557–566). MIT Press.

Manning, JR., Polyn, SM., Baltuch, G., Litt, B. Kahana, MJ. 2011. Oscillatory patterns in temporal lobe reveal context reinstatement during memory search Oscillatory patterns in temporal lobe reveal context reinstatement during memory search. Proceedings of the National Academy of Sciences, USA1083112893–12897.

MannEtal18Manning, JR., Zhu, X., Willke, TL., Ranganath, R., Stachenfeld, K., Hasson, U.Norman, KA. 2018. A probabilistic approach to discovering dynamic full-brain functional connectivity patterns A probabilistic approach to discovering dynamic full-brain functional connectivity patterns. NeuroImage180243–252.

Martin, J. Johnson, S. 2015. Target detection in visual search: unravelling the pupillary response Target detection in visual search: unravelling the pupillary response. Journal of Vision1512782.

McCorry, LK. 2007. Physiology of the autonomic nervous system Physiology of the autonomic nervous system. American Journal of Pharmaceutical Education7141–11.

McKinney, W. 2010. Data structures for statistical computing in Python Data structures for statistical computing in Python. Proceedings of the Python in Science Conference Proceedings of the Python in science conference (51–56).

Megreya, AM. Burton, AM. 2007. Hits and false positives in face matching: A familiarity-based dissociation Hits and false positives in face matching: A familiarity-based dissociation. Perception and Psychophysics6971175–1184.

Mill, RD., O'Connor, AR. Dobbins, IG. 2016. Pupil dilation during recognition memory: isolating unexpected recognition from judgment uncertainty Pupil dilation during recognition memory: isolating unexpected recognition from judgment uncertainty. Cognition15481–94.

Murdock, BB. 1962. The serial position effect of free recall The serial position effect of free recall. Journal of Experimental Psychology: General64482–488.

Naber, M., Frässle, S., Rutishauser, U. Einhäuser, W. 2013. Pupil size signals novelty and predicts later retrieval success for declarative memories of natural scenes Pupil size signals novelty and predicts later retrieval success for declarative memories of natural scenes. Journal of Vision1321–20.

Noudoost, B., Chang, MH., Steinmetz, NA. Moore, T. 2010. Top-down control of visual attention Top-down control of visual attention. Current Opinion in Neurobiology202183–190.

O'Craven, KM., Downing, PE. Kanwisher, N. 1999. fMRI evidence for objects as the units of attentional selection fMRI evidence for objects as the units of attentional selection. Nature.

Oliva, M. Anikin, A. 2018. Pupil dilation reflects the time course of emotion recognition in human vocalizations Pupil dilation reflects the time course of emotion recognition in human vocalizations. Scientific Reports848711–10.

Paller, KA. Wagner, AD. 2002. Observing the transformation of experience into memory Observing the transformation of experience into memory. Trends in Cognitive Sciences6293–102.

Paolacci, G., Chandler, J. Ipeirotis, PG. 2010. Running experiments on Amazon Mechanical Turk Running experiments on Amazon Mechanical Turk. Judgement and Decision Making55411–419.

Papesh, MH., Goldinger, SD. Hout, MC. 2012. Memory strength and specificity revealed by pupillometry Memory strength and specificity revealed by pupillometry. International Journal of Psychophysiology83156–64.

Park, M. Pillow, JW. 2012. Bayesian active learning with localized priors for fast receptive field characterization Bayesian active learning with localized priors for fast receptive field characterization. Advances in Neural Information Processing Systems Advances in neural information processing systems (2348–2356).

Pearson, K. 1901. On lines and planes of closest fit to systems of points in space On lines and planes of closest fit to systems of points in space. The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science2559–572.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O.Duchesnay, E. 2011. Scikit-learn: machine learning in Python Scikit-learn: machine learning in Python. Journal of Machine Learning Research122825–2830.

Peirce, JW., Gray, J., Simpson, S., MacAskill, MR., Höchenberger, R., Sogo, H.Lindeløv, J. 2019. PsychoPy2: experiments in behavior made easy PsychoPy2: experiments in behavior made easy. Behavior Research Methods511195–203.

Phillips, PJ., Wechsler, H., Huang, J. Rauss, PJ. 1998. The feret database and evaluation procedure for face-recognition algorithms The feret database and evaluation procedure for face-recognition algorithms. Image and Vision Computing165295–306.

Polyn, SM. Kahana, MJ. 2008. Memory search and the neural representation of context Memory search and the neural representation of context. Trends in Cognitive Sciences12124–30.

Posner, MI. 1980. Orienting of attention Orienting of attention. Quarterly Journal of Experimental Psychology3213–25.

Posner, MI., Walker, JA., Friedrich, FA. Rafal, RD. 1987. How do the parietal lobes direct covert attention How do the parietal lobes direct covert attention. Neuropsychologia251135–145.

Preuschoff, K., 't Hart, BM. Einhäuser, W. 2011. Pupil dilation signals surprise: evidence for noradrenaline's role in decision making Pupil dilation signals surprise: evidence for noradrenaline's role in decision making. Frontiers in Neuroscience51151–12.

Rabiner, LR. 1989. A tutorial on Hidden Markov Models and selected applications in speech recognition A tutorial on Hidden Markov Models and selected applications in speech recognition. Proceedings of the IEEE772257–286.

RangRitc12Ranganath, C. Ritchey, M. 2012. Two cortical systems for memory-guided behavior Two cortical systems for memory-guided behavior. Nature Reviews Neuroscience13713–726.

Rijn, HV., Dalenberg, JR., Borst, JP. Sprenger, SA. 2012. Pupil dilation co-varies with memory strength of individual traces in a delayed response paired-associate task Pupil dilation co-varies with memory strength of individual traces in a delayed response paired-associate task. PLoS One712e51134.

Roediger, HL. McDermott, KB. 1995. Creating false memories: Remembering words not presented in lists Creating false memories: Remembering words not presented in lists. Journal of Experimental Psychology: Learning, Memory, and Cognition21803–814.

Rundus, D. 1971. Analysis of rehearsal processes in free recall Analysis of rehearsal processes in free recall. Journal of Experimental Psychology: General89163–77.

Salinas, E. Thier, P. 2000. Gain modulation: A major computational principle of the central nervous system Gain modulation: A major computational principle of the central nervous system. Neuron27115–21.

Schneider, KA. 2011. Subcortical mechanisms of feature-based attention Subcortical mechanisms of feature-based attention. The Journal of Neuroscience31238643–8653.

Scho66Schönemann, P. 1966. A generalized solution of the orthogonal Procrustes problem A generalized solution of the orthogonal Procrustes problem. Psychometrika311–10.

Siegle, GJ., Steinhauer, SR., Carter, CS., Ramel, W. Thase, ME. 2003. Do the seconds turn into hours? Relationships between sustained pupil dilation in response to emotional information and self-reported rumination Do the seconds turn into hours? relationships between sustained pupil dilation in response to emotional information and self-reported rumination. Cognitive Therapy and Research273365–382.

Slooten, JCV., Jahfari, S., Knapen, T. Theeuwes, J. 2018. How pupil responses track value-based decision making during and after reinforcement learning How pupil

responses track value-based decision making during and after reinforcement learning. PLoS Computational Biology1411e1006632.

Soto, D. Blanco, MJ. 2004. Spatial attention and object-based attention: a comparison within a single task Spatial attention and object-based attention: a comparison within a single task. Vision Research44169–81.

Tan, L. Ward, G. 2000. A recency-based account of the primacy effect in free recall A recency-based account of the primacy effect in free recall. Journal of Experimental Psychology: Learning, Memory, and Cognition261589–1626.

Tan, L. Ward, G. 2008. Rehearsal in immediate serial recall Rehearsal in immediate serial recall. Psychonomic Bulletin and Review153535–542.

Tipping, ME. Bishop, CM. 1999. Probabilistic principal component analysis Probabilistic principal component analysis. Journal of Royal Statistical Society, Series B613611–622.

Torgerson, WS. 1958. Theory and methods of scaling Theory and methods of scaling. New York, NYWiley.

Treue, S. Trujillo, JCM. 1999. Feature-based attention influences motion processing gain in macaque visual cortex Feature-based attention influences motion processing gain in macaque visual cortex. Nature3996736575–579.

Tufte, ER. Graves-Morris, PR. 1983. The visual display of quantitative information The visual display of quantitative information (2) (9). Cheshire, CTGraphics Press.

Turk-Browne, N., Golomb, J. Chine, M. 2013. Complementary attentional components of successful memory encoding Complementary attentional components of successful memory encoding. NeuroImage.

Uddenberg, S., Newman, G. Scholl, B. 2016. Perceptual averaging of scientific data: implications of ensemble representations for the perception of patterns in graphs Perceptual averaging of scientific data: implications of ensemble representations for the perception of patterns in graphs. Journal of Vision16121081.

Uncapher, MR., Hutchinson, JB. Wagner, AD. 2011. Dissociable effects of top-down and bottom-up attentionduring episodic encoding Dissociable effects of top-down and bottom-up attentionduring episodic encoding. The Journal of Neuroscience.

TotalRecallUPenn Computational Memory Lab. 2015. Penn TotalRecall. Penn TotalRecall. Computer Software.

van der Linden, WJ. Glas, CA. 2000. Computerized adaptive testing: Theory and practice Computerized adaptive testing: Theory and practice. Springer.

van der Maaten, LJP. Hinton, GE. 2008. Visualizing High-Dimensional Data Using t-SNE Visualizing high-dimensional data using t-SNE. Journal of Machine Learning Research92579-2605.

van der Walt, S., Colbert, SC. Varoquaux, G. 2011. The NumPy Array: A Structure for Efficient Numerical Computation The NumPy array: A structure for efficient numerical computation. Computing in Science & Engineering1322–30.

Vartanian, O., Martindale, C. Kwiatkowski, J. 2007. Creative potential, attention, and speed of information processing Creative potential, attention, and speed of information processing. Personality and Individual Differences4361470–1480.

Vilberg, KL. Rugg, MD. 2008. Memory retrieval and the parietal cortex: A review of evidence from a dual-process perspective Memory retrieval and the parietal cortex: A review of evidence from a dual-process perspective. Neuropsychologia4671787–1799.

Võ, MLH., Jacobs, AM., Kuchinke, L., Hofmann, M., Conrad, M., Schacht, A. Hutzler, F. 2007. The coupling of emotion and cognition in the eye: introducing the pupil old/new effect The coupling of emotion and cognition in the eye: introducing the pupil old/new effect. Psychophysiology00130–140.

Wang, C., Huang, J., Brien, DC. Munoz, DP. 2020. Saliency and priority modulation in a pop-out paradigm: pupil size and microsaccades Saliency and priority modulation in a pop-out paradigm: pupil size and microsaccades. Biological Psychology153107901.

Waskom, M., Botvinnik, O., Okane, D., Hobson, P., David, Halchenko, Y.Lee, A. 2016. Seaborn: v0.7.1. Seaborn: v0.7.1.

Wittig, JH., Jang, AI., Cocjin, JB., Inati, SK. Zaghloul, KA. 2018. Attention improves memory by suppressing spiking-neuron activity in the human anterior temporal lobe Attention improves memory by suppressing spiking-neuron activity in the human anterior temporal lobe. Nature Neuroscience21808–810.

Xiao, J., Hays, J., Ehinger, K., Oliva, A. Torralba, A. 2010. SUN database: large-scale scene recognition from abbey to zoo SUN database: large-scale scene recognition from abbey to zoo. IEEE Conference on Computer Vision and Pattern Recognition. IEEE conference on computer vision and pattern recognition.

Yi, DJ., Kelley, TA., Marois, R. Chun, MM. 2006. Attentional modulation of repetition attenuation is anatomically dissociable for scenes and faces Attentional modulation of repetition attenuation is anatomically dissociable for scenes and faces. Brain Research.

Yonelinas, AP. 2002. The nature of recollection and familiarity: A review of 30 years of research The nature of recollection and familiarity: A review of 30 years of research. Journal of Memory and Language46441–517.

Zekveld, AA., Koelewijn, T. Kramer, SE. 2018. The pupil dilation response to auditory stimuli: current state of knowledge The pupil dilation response to auditory stimuli: current state of knowledge. Trends in Hearing22.

Ziman, K., Lee, MR., Martinez, AR., Adner, ED. Manning, JR. 2020. Feature-based and location-based volitional covert attention affect memory at different timescales Feature-based and location-based volitional covert attention affect memory at different timescales. PsyArXivdoi.org/10.31234/osf.io/2ps6e.