Exploring Visual Short-Term Memory in Ensembles via Change Detection

Athina Kakkos

This thesis is submitted in partial fulfilment of the Honours degree of Bachelor of Psychological Science (Honours)

School of Psychology

Faculty of Health and Medical Sciences

University of Adelaide

October 2019

Word Count: 9,482

Table of Contents

List of Figures

List of Tables

Abstract

Detecting changes in our visual environment is fundamental to our everyday functioning, for example navigating safely in traffic. Change detection is often quick and precise, with people able to capture meaning of a scene in under a second. However, people can fail to detect significant visual changes when a disruption occurs between views (Simons & Ambinder, 2005). It remains unclear just how much the finer visual details matter for detecting change in scenes. To explore whether people are more sensitive to changes in the summary, or details in natural scenes, we explored short-term visual working memory by manipulating change-size across two change detection experiments. Participants (*n* = 30) were presented with arrays and its summarised image average for 150ms with a blank 300ms mask in between. Confidence ratings for participant certainty in the change occurring was also investigated. Contrary to predictions, participants were better at discriminating changes in the summary statistic averages than detailed arrays. However, performance increased with proportion of change size and confidence as predicted. It was concluded change size and image representation does affect change detection, and more visual detail is not always necessary to detecting change - spare representations are not so sparse after all.

Declaration

This thesis contains no material which has been accepted for the award of any other degree of diploma in any University, and, to the best of my knowledge, this thesis contains no material previously published except where due reference is made. I give permission for the digital version of this thesis to be made available on the web, via the University of Adelaide's digital thesis repository, the Library Search and through web search engines, unless permission has been granted by the School to restrict access for a period of time.

Signature: Date: 01.10.2019

Acknowledgements

First and foremost, I would like to thank my supervisor, Doctor Rachel Searston. I am incredibly grateful to have learnt from such an enthusiastic, dedicated and generous mentor. You are not only a fabulous teacher but an inspiring person to work alongside. Your expertise and guidance in not only research (and coffee quality in Adelaide), but possible career paths post-honours has truly helped make this year one of the most rewarding, insightful and enjoyable. Without your wonderful patience and constructive feedback, I would not have been able to craft a thesis that I am proud of and enjoy telling people about – even when no one asks!

A special thank you to my family for your endless love and support; even when I said I was moving cities (and would not be able to attend many AFL Collingwood games). To my parents, thank you for your constant encouragement and taking the time to visit Adelaide. Your enthusiasm in helping me reach my goals and eagerness to take part in the research is greatly appreciated and does not go unnoticed. To my brothers, thank you for listening to my ideas and letting me know which were “coming up Millhouse”.

To the lab family, friends and 2019 honours cohort, thank you. From supportive discussions regarding university life (and lack of hot jam doughnuts in the city), to the delicious spread of study snacks and uplifting coffee runs; all have served as a comforting reminder that whilst this has been a demanding year, it has also been one filled with many brilliant laughs and great friendships. I wish you all the very best for the bright future ahead.

CHAPTER 1

Introduction

How well do we perceive the world around us? Classic studies on the limits of perception, memory, and attention suggest that our capacity to perceive every detail in our visual environment is severely impoverished (Cohen & Chun, 2017; Rensink, O'Regan & Clark, 1997; Simons & Ambinder, 2005). For example, in a study by Chabris, Weinberger, Fontaine & Simons (2011), a participant and experimenter jogged around a university campus where a mock fight with three confederates was planned in plain view. The fight involved shouting and grunting, which was only a few feet away from the participant and visible for at least 15 seconds. Only 35 per cent of participants reported seeing the fight during the lit night-time trial. The daytime trial found only 56 per cent of participants noticed the staged fight. A significant proportion of people managed to miss a loud and visible fight, with many surprised they did not notice it. In another real-world display of inattentional blindness, pedestrians were asked if they noticed a brightly coloured clown riding a unicycle near their walking path (Hyman, Boss, McKenzie & Caggiano, 2010). Of those using their mobile phone, only 25 per cent noticed the clown. These studies highlight that seemingly large or unusual changes to our visual environment go far more undetected than originally thought.

This research is contrary to the phenomenological experience that suggests we perceive an incredibly rich and detailed world from the moment we open our eyes (Cohen, Dennett, & Kanwisher, 2016b; Haun, Tononi, Koch & Tsuchiya, 2017). Can sparse visual representations give rise to rich visual experiences? In this thesis, I use a change detection task to explore what people can remember from sparse versus detailed views of the world.

1.1 Change Detection

Change detection is commonly used in visual perception to gauge human sensitivity to shifting stimuli. It refers to the visual processes involved in recognising alterations occurring in the world around us over time (Rensink, 2002). The ability to detect change is important in much of everyday life. Navigating road traffic, for example, involves a dynamic process of detecting and avoiding potential collisions. When a change is seen, attention is drawn to its location to facilitate further detailed visual processing (Pessoa & Ungerleider, 2004). Change detection tasks in the lab typically involve presenting observers with quick alterations (< 200 milliseconds) of a scene on a computer screen and commenting if a change occurred between alterations or not. These tasks are made challenging by the presence of mask, blank screen or implemented noise in between views to cover the quick alteration of a scene. This obscures any changes from online or real-time visual processing (Simons & Ambinder, 2005).

It is surprisingly difficult to detect changes from short-term memory even when we are deliberately searching for them (Brady, Konkle, Oliva, & Alvarez, 2009; Murphy & Murphy, 2018). Simons and Levin (1997) showed that 50 per cent of observers failed to notice that two people in a photograph had exchanged heads when shifting their eyes from one side of the photograph to the other. Changes introduced during eye movements or even in centreal vision can also go undetected (Rensink et al., 1997; Simons & Levin, 1997). Remarkably, people can fail to detect changes between two images separated by a blank screen even when the changing object is large (Cohen & Chun, 2017). This inability to detect changes due to a disruption or time-lag between views is often used to illustrate our surprisingly limited capacity to perceive and remember every detail of our visual environment.

One explanation for why people perform so poorly on change detection tasks is that the critical change (e.g., removing a sink from a kitchen scene) often preserves the summary statistics of the scene (e.g., the kitchen remains intact; Cohen et al., 2016a). People have a limited capacity to perceive particular details at a high resolution and are thought to rely on a summary of information across the visual field to resolve the entire scene. Disruptions to this summary information are much easier to detect because they change the 'gist' of the entire scene (Alvarez & Oliva, 2009; Brady & Alvarez, 2011). This summary information is thought to be represented as an ensemble or an average. If so, then people may display a sensitivity to changes in the average scene given the absence of particular items. In my honours project, I test people's sensitivity to changing ensembles of scenes and paintings, represented as averages or arrays of images.

First, I look to research on gist perception, ensemble representations, and studies of visual short-term working memory, as a guide to understanding failures of change detection and how people can be *looking* but not always *seeing* significant objects that come into their field of view.

1.2 Gist Perception

Gist perception refers to the amount of perceptual information an observer can comprehend within a glance of a scene (Oliva & Torralba, 2006). People can capture a scene or objects' general meaning, orientation, size and shape—it's gist—within a brief glance (Howe, 2017; Koehler & Eckstein, 2017; Oliva & Torralba, 2006). Humans can also recognise the general substance of even blurry images when presented for just 100 to 200 milliseconds (Schyns & Oliva, 1994). The gist of a scene can be extracted with an accuracy level of over ninety per cent in some tasks (Rousselet, Joubert, & Fabre-Thorpe, 2005), and significantly above chance even when the images have been reduced down to a handful of pixels (Searston, Thompson, Vokey, French, & Tangen, 2019). Experts in various fields also display an increased ability to extract the gist of complex images quickly and accurately. For example, fingerprint experts can detect prints belonging to the same person but left by different fingers (Searston & Tangen, 2017) and they can identify prints presented in noise with little time (Thompson & Tangen, 2014). Remarkably, Brennan et al. (2018) found that radiologists can detect abnormalities in mammograms at above-chance levels after a momentary glimpse of an image containing an abnormality. Similarly, others who have investigated gist perception in abstract artworks have found that exposure times as short as 50ms can be enough for people to form a judgment of aesthetic (Schwabe, Menzel, Mullin, Wagemans & Redies, 2018).

Gist perception is thought to guide the allocation of attention and use past knowledge to decipher the more detailed elements in our visual environment (Sampanes, Tseng & Bridgeman, 2008; Torralba, 2009). For example, a brief glimpse of a kitchen can be used to direct your eyes to likely locations of the sink as gist comprises the spatial layout of the room and helps to constrain the location of particular objects. Research exploring gist perception has also distinguished between *global* and *local* features within a scene (Navon, 1977; Rousselet et al., 2005). Global features capture the holistic structure of a scene, with many shape and texture descriptors falling under this category. Local features, on the other hand, are computed at multiple key points within a scene and thus are more robust to occlusion or clutter in the scene than global perception (Franconeri, Alvarez & Enns, 2007). Global features are thought to be perceived first, in the time course of perceptual experience, guiding further local level feature analysis within a scene (Navon, 1977). Summary statistics, such as the average hue, luminance or spatial frequency of a scene, are also thought to be central to gist perception (Oliva & Torralba, 2006), and these summary statistics are referred to as ensembles (Jackson-Nielsen, Cohen & Pitts, 2017).

1.3 Ensemble Representations

Ensemble coding or ensemble representation is the idea that the visual system, rather unconsciously, represents multiple items seen into a single, average depiction, therefore creating a single summary statistic (Cohen et al., 2016a). Ensemble coding is the ability to extract summary statistical information from groups of similar objects and is thought to be useful for organising information in working memory and understanding the gist of a visual scene (Jackson-Nielsen et al., 2017; Whitney & Yamanashi-Leib, 2018). Summary statistics can be represented across a range of visual dimensions including average orientation, position and facial expression (Alvarez & Oliva, 2008; Whitney, Haberman & Sweeny, 2014). For example, we are able to perceive a lawn without viewing every single blade of grass. That is, ensemble perception enables us to see past redundant features to the statistical summary or average of the scene. Such ensemble or summary representations are thought to drive the compelling impression that we perceive a complete and accurate picture of the visual world (Noe, Pessoa, & Thompson, 2000; Haberman & Whitney, 2009). Thus, it is thought to serve as a cognitively economical driver in perceiving the gist of a scene.

1.4 Visual Short-Term Memory

Visual short-term memory (VSTM) research suggests that people's ability to remember lists or arrays of specific items has a critical limit (Cowan, 2001; Miller, 1994). This limit appears to increase via a process of ensemble coding or averaging, which effectively reduces an array of items to a single summary statistical representation.

Notably, as highlighted by Bateman, Ngiam & Birney (2018) there is a distinction between VSTM and visual working memory (WM). VSTM is the cognitive system's process of rapidly creating representations of visual information and actively preserving it for a few seconds to aid the requirements of ongoing activities (Luck & Vogel, 2013). Meanwhile, WM research involves brief exposure times, often less than one second to assess encoding performance (much similar to VSTM research) but involves the manipulation of information as opposed to only retaining.

Research into the VSTM system highlights a limited capacity of only holding 3-4 items, or ‘slots’ (Cowan, 2001; Luck & Vogel, 1997), as opposed to long-term memory that can store thousands of items (Brady, Konkle & Alvarez, 2011). Visual perception research approaches the domain of memory by attempting to determine *what* is being represented through change detection tasks. This aims to better understand the architecture of visual cognition and its capacity; often by exploring its limits via change blindness tasks (Suchow, Fougnie, Brady & Alvarez, 2014). Whilst there are a fixed number of capacities, or ‘slots’, more studies have shown 7-8 items may be attended to in visual attention and working memory from the traditional 3-4 items (Franconeri et al., 2007; Howe, Cohen, Pinto & Horowitz, 2010; Miller, 1994).

An explanation for this increase in ‘slots’ is *relational grouping*. This has been shown to enhance recall in change detection tasks due to the encoding of configural relationships between objects (that is, recording the associations between emergent features) amongst certain traits in the scene (Rensink, 2000; Bateman et al., 2018). For example, in face perception, humans typically detect two eyes, a nose and mouth (first-order relations) and then process this information holistically by ‘gluing’ the features together to then assist in determining they are indeed viewing a face (Maurer, Le Grand & Mondloch, 2002). This advocates that relational grouping further assists the retention of individual in memory (Jiang, Olson & Chun, 2000).

Whilst Cohen et al. (2016b) contend even 7-8 slots do not sufficiently lend strength to the debate of perceptual experience nor explain its richness; Brady and Tenenbaum (2013) calculated human memory capacity via a Bayesian model to account for the assumption that observers remember not just individual items but also a summary of the display. This was highlighted via a change detection task with stimuli that consisted of 5 ×5 coloured dots. When displays had limited structure, memory capacity was around 4.5 items; although, when more structure was added, an estimated 24-25 items were held in memory, suggesting observers encode a few individual items alongside a visual summary of the whole display.

Despite the assumption that ensemble representations are thought to comprise the average of a scene's component parts, we know little about people's ability to detect changes to visual averages (i.e., sum total divided by the number of items). Therefore, in my thesis, I examine the extent to which people can detect changes in average representations of natural images, such as scenes and paintings. My broad aim is to provide further insight into the nature of visual short-term working memory for complex image sets.

1.5 The Current Project

Prior work demonstrates that people can detect changes in simple visual displays shown on-screen between 130 to 350 milliseconds, with natural scenes processed faster than artificial ones, and gist accuracy of around 90 per cent (Rousselet et al., 2005; Schwabe et al., 2018). VSTM and change detection is commonly explored by using simple displays of randomly arranged stimuli; believed to prevent contamination from memory systems and familiarity (Murphy & Murphy, 2018). However, less is known about how sensitive people are to changes in natural scenes and images. The current project involves two twin experiments exploring people's ability to detect global changes in the average of natural scenes and images. In particular, I test people's sensitivity to change in the presence or absence of local image detail by presented a series of natural images simultaneously as an array (see Figure 1A) or as an average (see Figure 1B). That is, are people better at remembering and detecting changes to the kitchen without the sink?

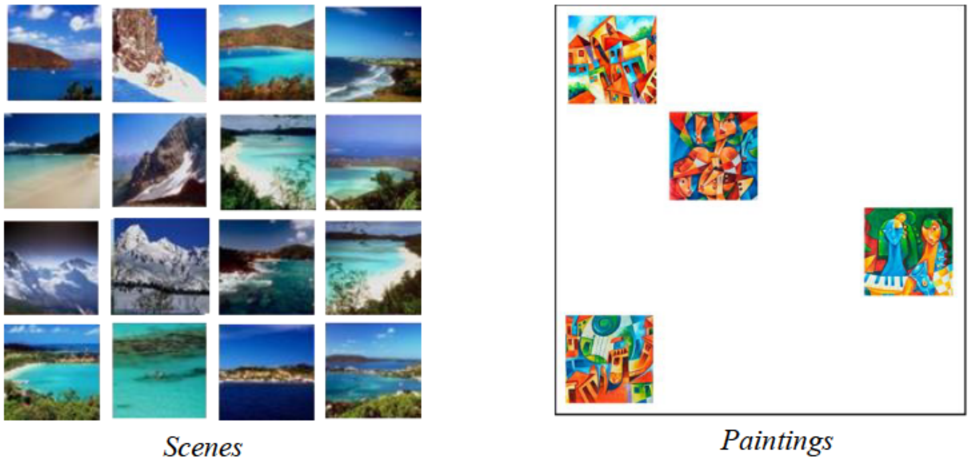


Figure 1A. Example of Arrays used in Experiment 1 from the Scenes domain (natural landscapes) and Experiment 2 of the Paintings domain (Cubist style).

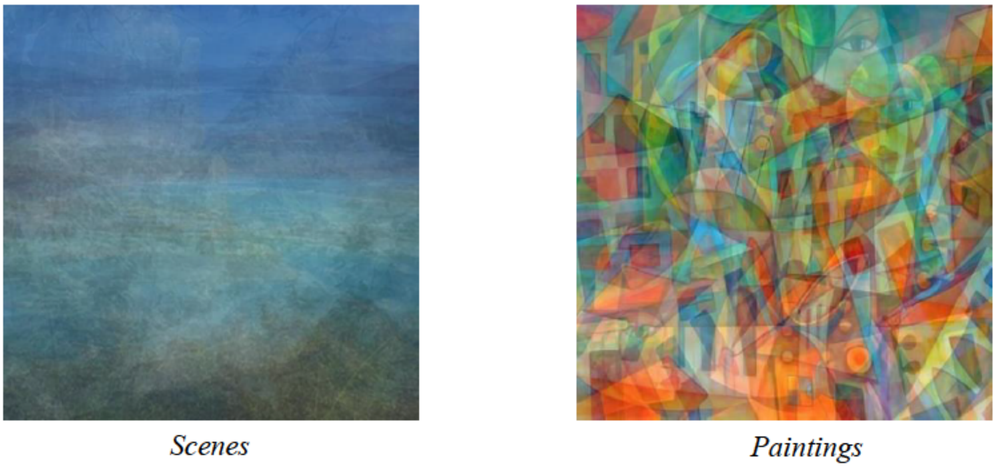


Figure 1B. Example of Averages used in Experiment 1 from the Scenes domain (natural landscapes) and Experiment 2 of the Paintings domain (Cubist style).

As mentioned above, one reason why people may perform poorly in change detection tasks is that the changes are often made locally to objects within the visual scene (e.g., a potted plant appearing and disappearing) without disrupting the global summary statistics (e.g., colour, luminance; Cohen et al., 2016b). Yet, Brady and Tenenbaum (2013) have suggested observers not only create a visual summary of the whole display, in line with Cohen et al. (2016b) stance on sparse perceptual experience, but encode a set of individual objects alongside the summary statistic when computing the overall gist.

The two experiments I present below were designed to tease apart these two opposing explanations for failures of change detection. I expected that participants would be more sensitive to changes in the arrays than the averages as these retain the local properties of the images, plus the global properties of the display. In the arrays condition, the location of the images are left intact across views, so as the images change, the global ensemble of the images also changes. The extent that participants are relying on the local detail in the images to detect changes, above and beyond the global ensemble of the array, proposes their performance should be higher with the arrays than the averages. On the other hand, if ensemble representations resembling the statistical average of a scene are critical to the perception of changes in a visual display, reducing the images to an average may boost change detection performance, even in the absence of local features. Based on the above line of thinking I predicted the following results:

1. Participants will be more sensitive to changes as the size of the change increases.
2. Participants will be more sensitive to changes in the arrays than the averages.
3. Participants will be more confident as the size of the change increases.
4. Participants will be more confident with the arrays compared with the averages.

My predictions were identical for Experiment 1 and 2, with the exception that change size was operationalised as the number of changed images in my first experiment (i.e., 1, 2, 4 or 8 images changed out of 16 images), and number of total images in my second experiment (i.e., 1 image changed out of 2, 4, 8, or 16 images). Both experiments were conducted concurrently with the same sample of participants to see if the pattern of results would replicate across different variants of the same manipulation.

CHAPTER 2

Experiment 1: Change Alteration

In Experiment 1, I explore people's sensitivity to change within averages (i.e. global change) and arrays (i.e. local change) of Natural and Urban scenes, and Cubist and Impressionist paintings. As I was not interested in short term memory for scenes or paintings specifically, I did not include stimuli as an explicit factor in my design and planned to average participants' responses for my main analyses for generality. A gap-contingent technique was used to investigate change detection, which involves a mask or blank field in between the original and changed stimulus (Rensink, 2002). I also used a one-shot change detection approach, further outlined by Rensink (2002), whereby the change is made once during each trial and does not 'flicker' between the mask and change stimulus. This one-shot task is designed to minimise the involvement of eye movements and long-term memory in change detection, enabling me to focus my analysis on VSTM for the averages versus the arrays.

2.1 Method

I used a 2 (change: no change, change) × 4 (change size: 1, 2, 4, 8 images changed out of 16) x 2 (representation: arrays, averages) fully within-subjects design, with representation as the key manipulation for investigating people's ability to detect global versus local changes. I included the change size manipulation to better isolate participants' capacity for detecting changes with the arrays and averages. Half of the trials consisted of arrays with 16 images, and the other half comprised the averages of those 16 images. The images in the arrays and averages were paired so that each average consisted of the exact same 16 images as one of the array trials, but the arrays of images were randomly sampled for each participant, so the arrays (see Figure 2 for an example) and averages (see Figure 3) consist of different samples of images on each new trial sequence.

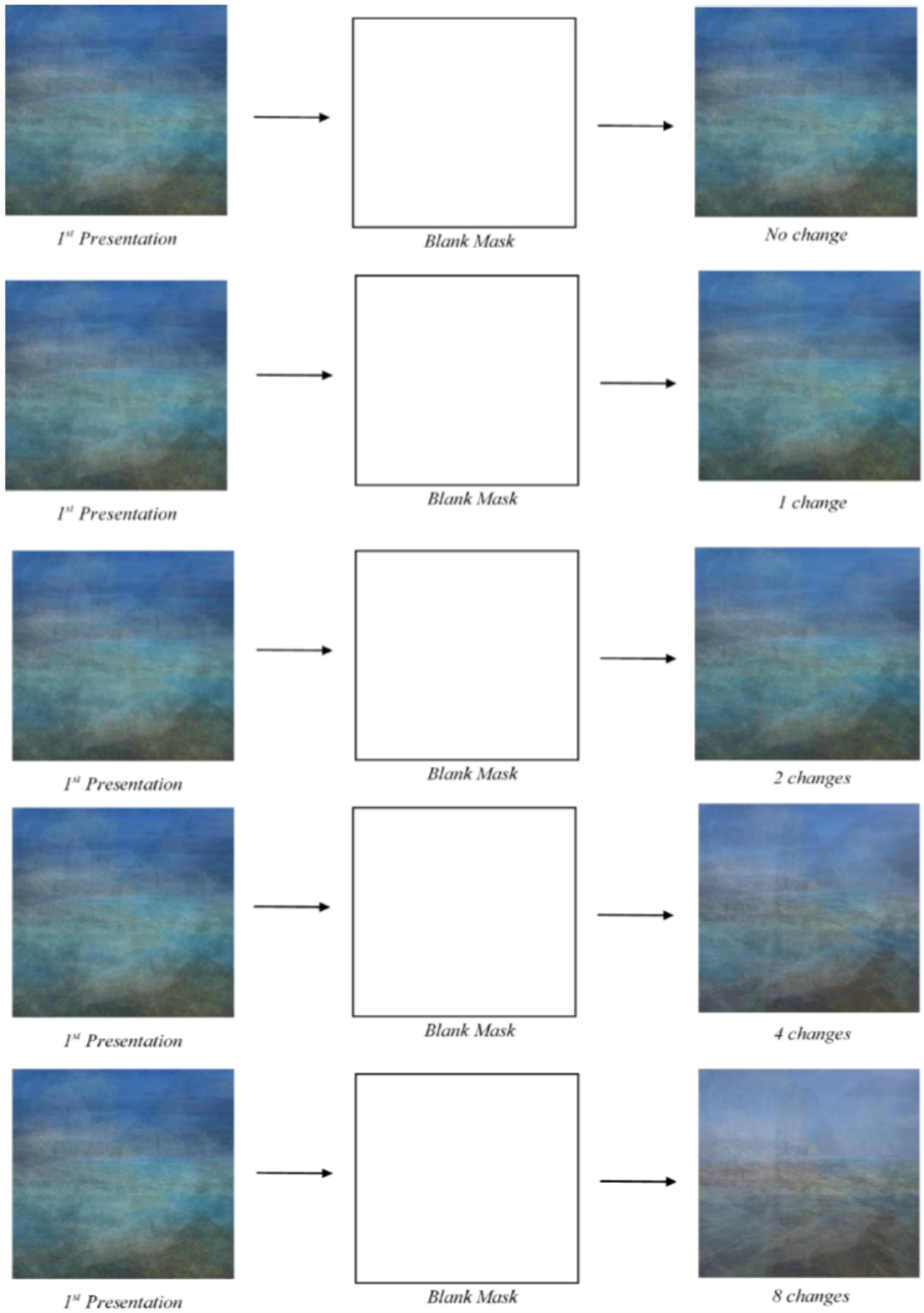


Figure 2. Example of Averages from the Scenes domain (natural landscapes) showcasing the trial sequence with altering number of change conditions. First presentation is displayed for 150ms, followed by blank mask for 300ms and then second presentation with either none or 1, 2, 4, 8 changes for 150ms.

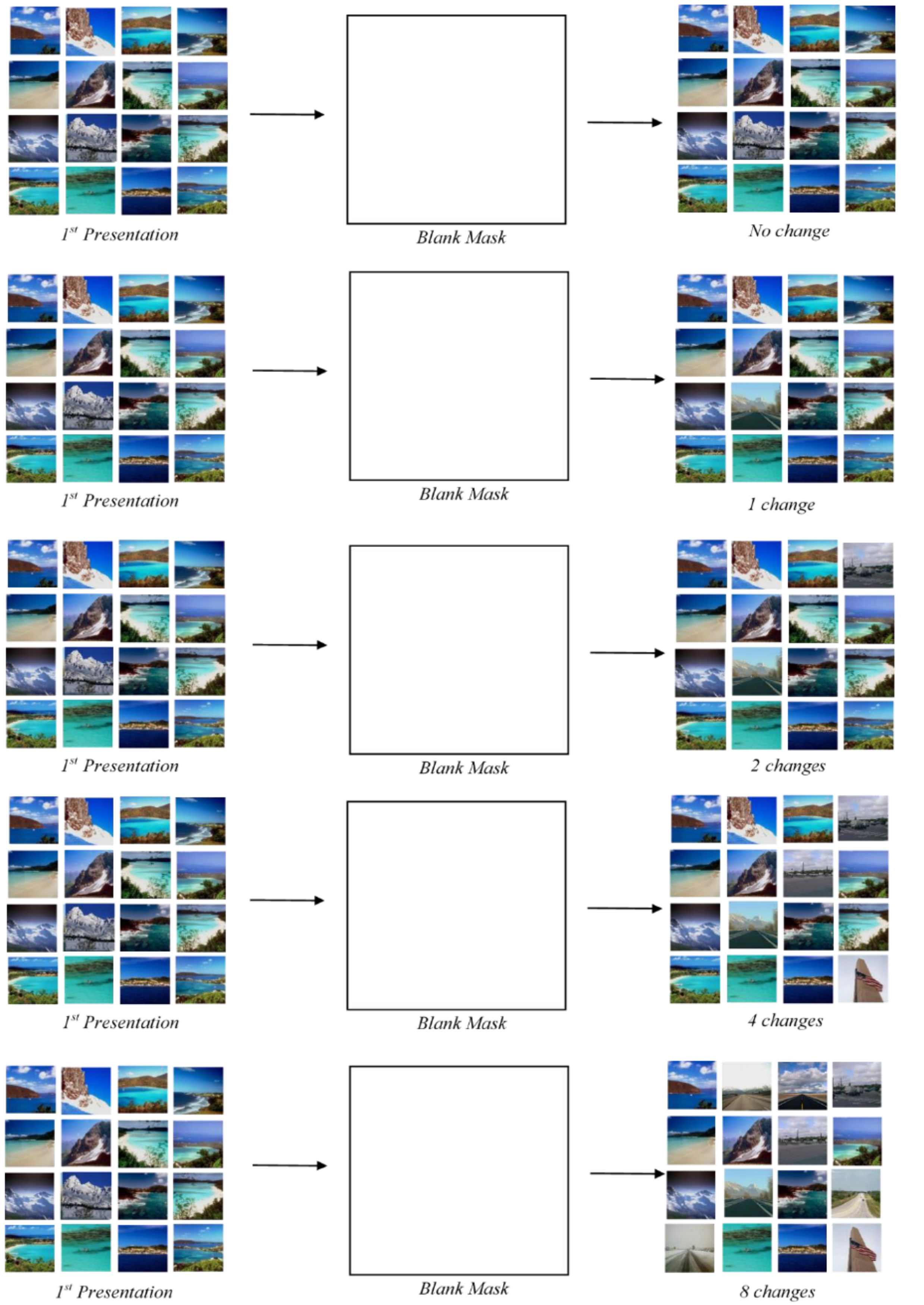


Figure 3. Example of Arrays from the Scenes domain (natural landscapes) showcasing the different change size conditions. First presentation displayed for 150ms, followed by blank mask for 300ms and then second presentation with either no or 1, 2, 4, 8 changes for 150ms.

Within the blocks of arrays and averages, a random half of trials were “No Change” trials, where none of the images changed between views, and the other half were “Change” trials, where either 1, 2, 4, or 8 images were swapped out for images from the opposite category. For example, on a change trial consisting of an array of 16 natural scenes in the 1-change condition, a random natural scene image would be swapped out for an urban scene image in the second presentation of the array (e.g., first glance includes 16 natural scenes, second glance includes 15 natural scenes and 1 urban scene). Likewise, in the 2-changes condition, two random natural scene images would be swapped for two urban scenes in the array, four swaps in the 4-change condition, and eight swaps in 8-change condition. Similarly, 1, 2, 4, or 8 Cubist paintings were swapped for Impressionist paintings and vice versa for paintings trials. An average image was generated for every set of original 16 images and every set of changed 16 images, so that each changed and unchanged array in every change size condition could be repeated in the average condition for each participant. The blocks of averages and arrays were counterbalanced across participants to control for order effects.

2.1.1 Materials

***2.1.1.1 Painting*s.** The paintings are a subsample from the ‘How Low Can You Go?’ collection (Searston et al., 2019). The full collection contains 5,184 paintings, made up of 18 different paintings by 72 different artists, in each of four different artistic styles (Cubism, Impressionism, Realism, and Renaissance; 288 artists in total, available at https://osf.io/kuja8/). All of the paintings in the ‘How Low Can You Go?’ collection were originally cropped to the centre of the shortest dimension using a 1:1 (square) aspect ratio, resized using nearest neighbour scaling to 800×800 pixels, and converted to jpeg format. All signatures had also been removed using the “Content Aware” fill tool in Photoshop. I used the 1,296 Cubist and 1,296 Impressionist paintings in the current experiment (2,592 painting in total).

2.1.1.2 Scenes. The scenes were a subsample of urban and natural scenes from the LabelMe dataset (Oliva & Torralba, 2001; available at http://cvcl.mit.edu/database.htm). The Natural category contains 359 beach, 328 forest, 374 mountain, and 411 open country scenes (1,472 natural scenes in total). The Urban category contains 260 highway, 308 inside city, 292 street, and 356 tall building scenes (1,216 urban scenes in total). All of the images were originally coloured, in jpeg format, and were standardised at 800 × 800 pixels.

2.1.1.3 Arrays. My supervisor developed a computer application that automatically sorted these two image sets into 80 Impressionist, Cubist, Natural and Urban subfolders. Each of the subfolders contains a different random sample of 16 images from that category that served as the original unchanged arrays (80×4 = 320 subfolders of 16 images in total). These subfolders were then duplicated four times to generated 4 additional sets of 320 subfolders for the four change size conditions: 1, 2, 4, or 8 images changed out of 16. My supervisor developed separate computer application that automatically 'swapped' 1, 2, 4 or 8 images in each subfolder with 1, 2, 4, or 8 images from the subfolders of the opposite category (e.g., Impressionists for Cubists, Natural for Urban), within each of change size folders. In total this process produced five folders labelled "no-change," "1-change," "2-changes," "4-changes," and "8-changes." Within each of those folders, there was a separate folder for "cubist," "impressionist," "natural," and "urban" arrays and within each of those subfolders there were 80 further subfolders containing 16 images—resulting in a total of 1600 arrays to sample from, or 80 arrays per category and condition.

2.1.1.4 Averages. A separate average image was generated for all 1600 arrays and these were embedded within the same file structure. That is, each of the 1600 subfolders contained the 16 original images and one average of those 16 images. I generated a subsample of 320 averages using an action in Adobe Photoshop (CC 2019) to group and summarise the arrays as a pilot of different aggregation methods. The final full set of 1600 averages were created using a MatLab (R2018b) script that loops through each of the folders, combines all of the images in each folder into a single pixel matrix, and then averages them by adjusting the opacity of each to 1/x, where x is the number of images in the average (16 in this experiment). In other words, the arrays preserve all original details in each set of 16 images, while the averages are a summary of those same 16 images.

***2.1.1.5 Presentation*.** The arrays were presented centrally on the computer screen in grids of 16 images at a reduced size of 200×200 pixels: a total array size of 800×800 pixels, plus 11 pixels of image separation. The averages were presented centrally on the computer screen at 800×800 pixels. I equated the image size of the averages with the total size of the image arrays (rather than equating the averages with the size of the individual images in the arrays) so that the stimuli were similarly distributed across the visual field in each condition.

2.1.2 Participants. A total of 30 participants were recruited (15 female, 15 male) from the general public via recruitment flyers placed around campus and social media, word of mouth, and The University of Adelaide's School of Psychology SONA Research Participation System (RPS, see Appendices A-D for recruitment poster, SONA advertisement, information sheet and consent form). All participants were over the age of 18 (M = 25.17, SD = 9.23) with normal or corrected-to-normal vision and compensated with a $20 Coles-Myer gift-card, although eight participants generously opted to volunteer their time in lieu of a gift-card. First year students recruited via the RPS were awarded one course credit for their time. Every participant gave informed consent and experimental procedures were approved by the School of Psychology Human Research Ethics Subcommittee at the University of Adelaide (approval number H-2019-73).

2.1.3 Procedure. Participants first read an information sheet about the project, completed a consent form and then began the experiment on a 13-inch MacBook Pro laptop (the application for presenting the experiment was developed in LiveCode Community). The first screen asked participants to input demographic information that was used to generate a unique, de-identifiable code for their data. The second screen then displayed an instructional video demonstrating the change detection task. The video included several examples of the task, the stimulus sets, each condition,, and an explanation on how to use the response scale (Experiment 1 video available: https://www.youtube.com/watch?v=3kfsb6jq980). After the video concluded, participants pressed a button to begin the change detection task.

On each trial, participants were presented with either an array of 16 images organised as a 4×4 grid, or an average of those 16 images summarised into one overall image for 150 milliseconds. This first presentation was followed by a 300-millisecond blank mask and then a second presentation of the same array or average for 150 milliseconds. The second presentation either had no changes (on the no change trials) or included either one, two, four or eight changed images from on change trials as described above. A response scale appeared in the centre of the screen immediately after the second presentation (see Figure 4). Participants were asked to indicate how strongly they believed a change occurred or not on a 12-point, forced choice, confidence scale; “no change” responses ranged from 1 (sure no change) to 6 (unsure no change) and “change” responses ranged from 7 (unsure change) to 12 (sure change). The scale remained on-screen until a selection was made, but if participants took longer than four seconds to respond they were prompted to try and respond within this time on future trials. Participants completed 256 trials in total, including 128 paintings and 128 scenes trials. Within each block of 128 trials, 64 consisted of ‘array’ trials and 64 consisted ‘average’, with a random half of participants viewing the paintings first and the other half viewing the scenes first. Each block of 64 trials was further divided into the four change size conditions (1, 2, 4, or 8 images changed), with a random half being no change trials (i.e., the images remained the same in both presentations) and the other half being change trials (i.e., one or more images changed in the second presentation). All of the trials within each of the counterbalanced arrays and averages blocks were presented in a different random order to each participant. In other words, the average and array trials are presented in two separate blocks within two separate paintings and scenes blocks, and are counterbalanced across participants. All other trials are randomised at the level of the individual participant.

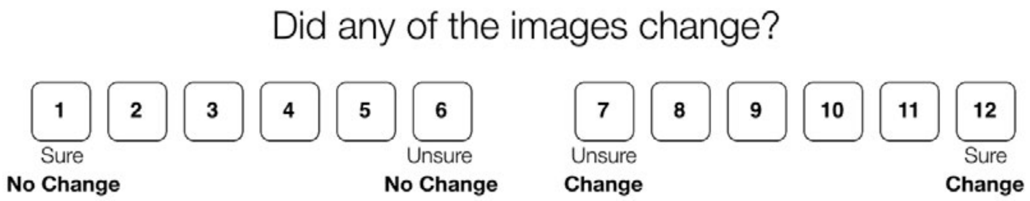


Figure 4. Example of forced-choice response scale shown after both image presentations.

2.2 Results

I investigated how the number of changes within each representation type (arrays and averages) affected participants' change detection ability by measuring their discriminability and confidence across conditions. First, as a basic measure of performance, each participants' proportion correct was computed by averaging the number of correct responses they made across all trials within each condition. For the main analyses comparing participants' ability to detect changes in the averages and arrays across change size conditions (1, 2, 4, 8 images changed), each participants' empirical area under the ROC curve (W or AUC) was computed based on their cumulative confidence ratings on the change and no change trials (see Hanley & McNeil, 1982 and Vokey, 2016 for the method used to compute AUC). AUC summarises participants' hit (say "change" on true change trials) and false alarm (say "change" on true no change trials), while also accounting for their confidence. An AUC score of 1 suggest perfect change detection, while an AUC of .5 indicates chance change detection.

Each participants' average confidence was also computed by converted their ratings on the 12-point forced-choice scale to a score out of 6, where a confidence rating of "1" indicated the participant was unsure and 6 indicated they were “Sure” in their rating. Ratings of 1 and 12 on the original scale were converted to confidence scores of 6 (“Sure”), ratings of 2 and 11 to confidence scores of 5, and so on, with ratings of 6 and 7 on the original scale representing a confidence score of 1 (“Unsure”). These collapse confidence scores were then averaged for each participant.

All analyses were performed in R (version 3.5.2), using RStudio (version 1.2.1335) and R Markdown (version 1.14) for documentation (see Appendix E-J for the code and output associated with all reported analyses). All plots included in this thesis were produced with the ggplot2 package (version 3.2.1). As Experiment 1 and 2 share similar manipulations of change size and representation, serving as close replications of one another, the interpretation of results will be left to the general discussion, in Chapter 4 of the thesis.

2.2.1 Descriptives. Overall, participants performed better in the averages (M = .65, SD = .16) than the arrays (M = .59, SD = .18). Table 1 shows participant sensitivity and proportion correct in each condition. On visual inspection of the data (see Figure 6), there were a few outlying observations, however these were retained in the dataset as none met the data exclusion criteria. As the data in the average and array conditions appeared to be normally distributed (see Figure 5A), parametric analyses were carried out as planned in the preregistration. Mauchley’s sphericity tests were performed for each analysis that included change size as a factor (which had more than two levels) and appropriate corrections were applied to the reported p values where this assumption was violated. more conservative statistics implemented where sphericity is violated.

To check for floor effects, a paired t-test was conducted comparing participants’ AUC scores to the simulated randomly responding participants. Participants were indeed performing significantly above chance; $t(239) = -9.397$, $p < .001$, 95% CI [-0.179, -0.117]. This result suggests that an overall floor effect was not present.

2.2.2 Discriminability. Discriminability was computed as Area Under the Receiver Operating Characteristic (ROC) Curve (AUC). AUC provides an indication of the extent to which a stimulus is accepted as a member of a particular target category, in this case change or no change categories (Vokey, 2016). AUC is a frequently used analysis in signal detection experiments which employ two-alternative forced choice techniques—in this case ‘change’ or ‘no change’ (Hanley & McNeil, 1982). The proportion of ‘hits’ (correct identification of a condition) are mapped against the proportion of false alarms (incorrect identification of a target). Discriminability asks how well participants can distinguish change trials from no change trials with 0.5 indicating at-chance performance and 1 equating to perfect performance, thus high sensitivity to distinguishing between hit and false alarm trials.

As shown in Table 1, participants’ discriminability of change and no change trials with the arrays and averages increased alongside the increasing number of changes. In the 1-change condition, participants performed close to chance with the arrays and averages, similar to flipping a coin. However, as the change size increased, so did participants’ sensitivity to detecting changes across the board. This trend lends support to Prediction 1. Notably, a change size of 8 images out of 16 produced the largest difference in sensitivity (10 per cent) between the averages and the arrays, suggesting that the average advantage might be strongest with larger changes. Similar to discriminability, as change size increased, the proportion of correct responses also increased, with averages outperforming arrays, and the 8-change condition again producing the largest difference.

Table 1

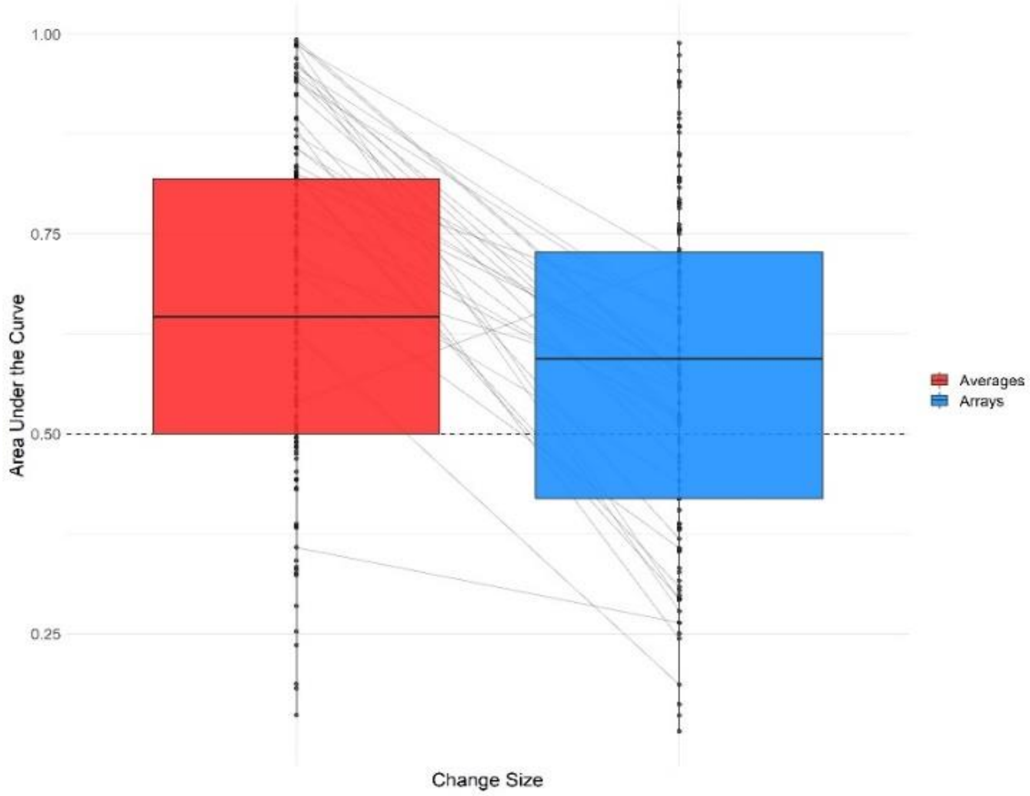
Experiment 1 Discriminability and Proportion of Correct Scores across Representation and Number of Changes

	Discriminability (AUC)		Proportion Correct	
	Averages	Arrays	Averages	Arrays
Number of Changes	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)
1 Change	.47 (.14)	.49 (.15)	.57 (.08)	.60 (.10)
2 Changes	.58 (.16)	.50 (.171)	.65 (.09)	.61 (.11)
4 Changes	.68 (.17)	.61 (.20)	.72 (.09)	.68 (.12)
8 Changes	.84 (.14)	.74 (.19)	.81 (.10)	.76 (.12)

Note. The AUC scores refer to mean (and standard deviation) Area Under the Curve for each condition in experiment.

After inspecting and visualising the data, I conducted repeated-measures ANOVAs on participants AUC scores with representation (arrays vs. averages) and change size (1, 2, 4, 8 images changed) as a within-subject factors. Mauchly's test indicated that the assumption of sphericity was violated for the change size condition, $W = 0.65$, $p = .035$, indicating significant heteroscedasticity or unequal variability in AUC scores across change size conditions. To minimize the risk of increasing in Type 1 error, the more conservative Greenhouse-Geiser (GG) corrected p value is reported for change size, $\varepsilon = .76$, $p < .001$, (Field, Miles & Field, 2012). There was a significant effect of representation on change detection with a small-medium effect size (see Bakeman, 2005 for analysis of generalised eta-squared effect size conventions in ANOVA) , $F(1, 29) = 6.48$, $p = .02$, $\eta^2_G = .03$ (see Figure 5A). That is, people were better at detecting changes in the averages compared with the arrays. Additionally, there was also a significant effect of change size change detection with a large effect size, $F(3, 87) = 108.10$, p[GG] $< .001$, $\eta^2_G = .34$ (see Figure 5B).

A.



B.

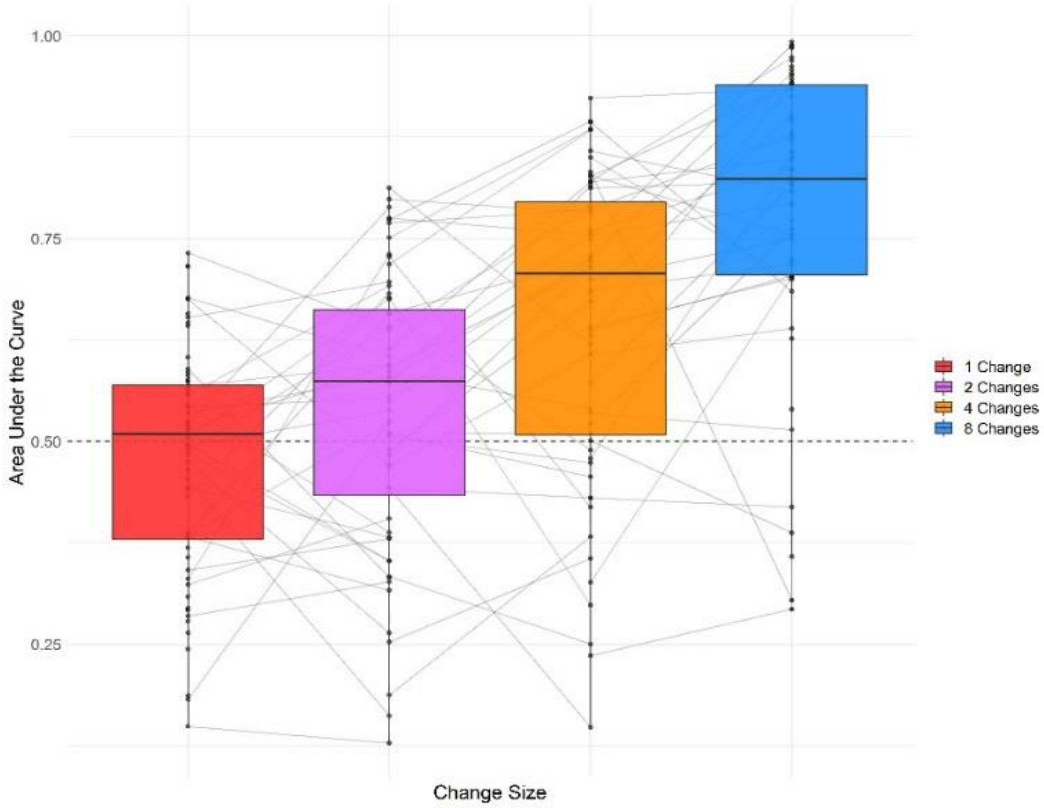


Figure 5. Panel depicts (A) main effect of participants' discriminability in performance across both representations, collapsed across number of changes, and (B) main effect of discriminability in sensitivity to the number of changes, collapsed across representation (arrays and averages).

Moreover, a significant but small interaction between representation and change size was also found, $F(3, 87) = 4.35$, $p = .01$, $\eta^2_G = .02$ (see Figure 6). Polynomial contrasts further revealed a significant linear trend over change size for arrays ($p < .001$) and averages ($p < .001$). That is, greater sensitivity was observed at with larger change sizes (i.e., number of images changed).

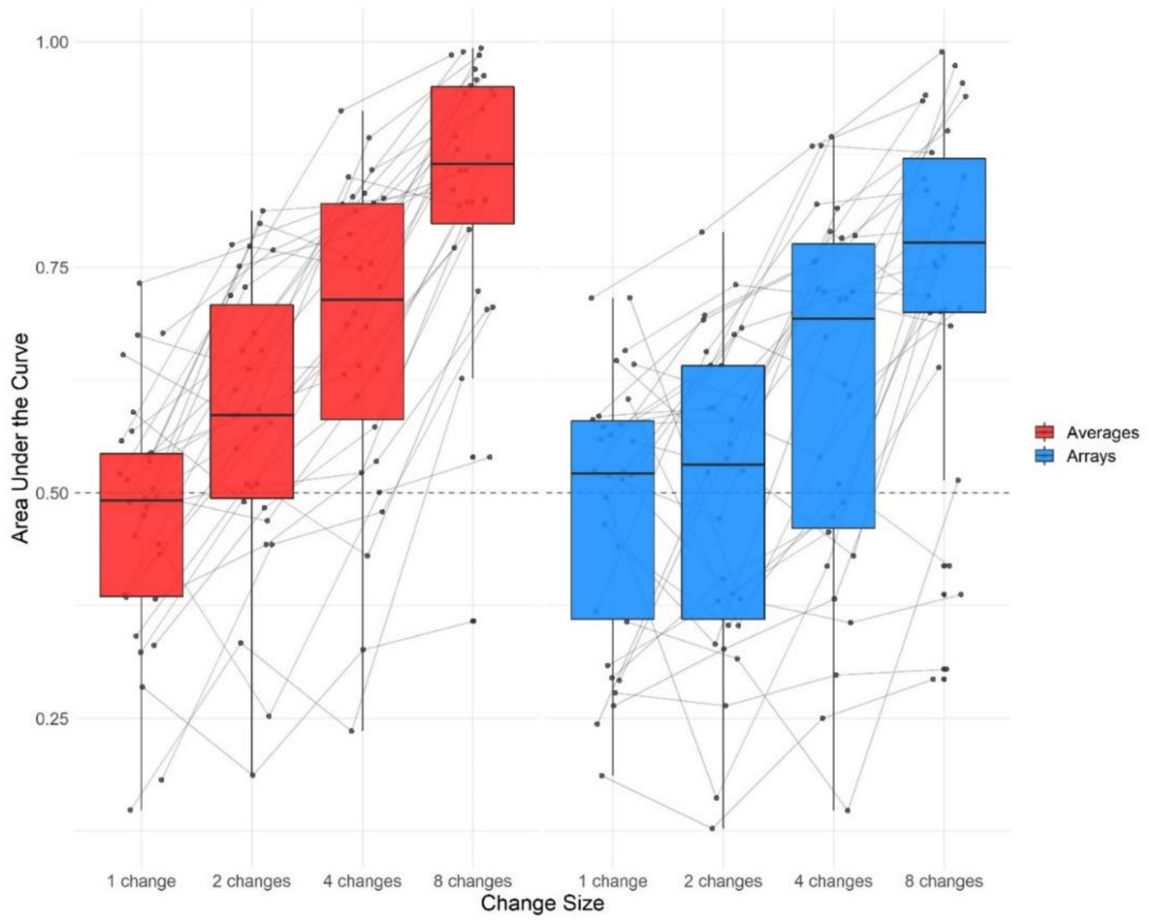


Figure 6. Panel depicts interaction effect between representation (Averages and Arrays) and change size (number of changes)

2.2.3 Confidence. To examine participants' if participants' confidence in their change detection ability changed across conditions, I subjected their mean confidence scores (out of 6) to the same set of analyses as their AUC scores. Capturing participant confidence ratings also assisted in computing their sensitivity (Massoni, Gajdos, Vergnaud, 2014). Participants were more confident at detecting changes in the averages (M = 3.33, SD = 1.44) than the arrays (M = 2.86, SD = 1.29) , and were most confident when there were eight changed images (M = 3.77, SD = 1.38) compared to the other smaller change size conditions (see Figure 7). Confidence in the 1-change and 2-change size conditions within arrays were similar: M = 2.74 (SD = 1.32) and M = 2.71 (SD = 1.27) respectively. Additionally, confidence in detecting changes with the arrays in the 8-change condition (M = 3.13, SD = 1.31) was similar to the 2-change condition for the averages (M = 3.14, SD = 1.31).

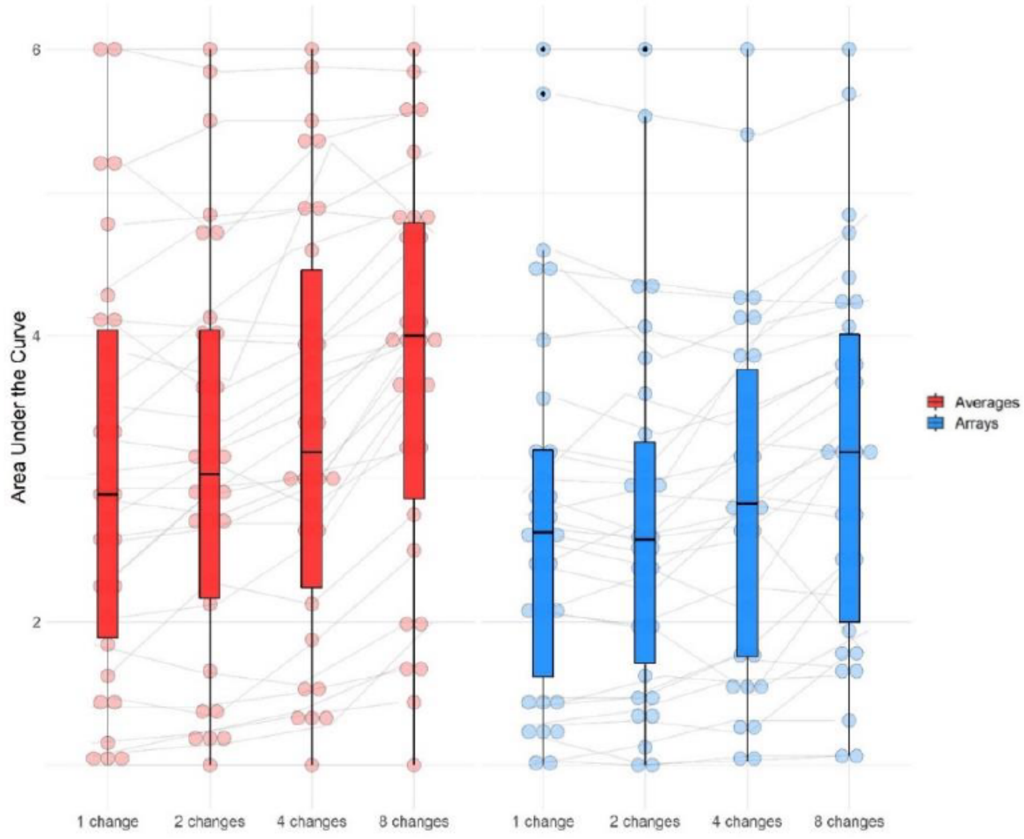


Figure 7. Confidence across change size in Experiment 1

A within-subjects ANOVA revealed a significant main effect of representation on confidence with a small-medium effect size, $F(1, 29) = 17.52, p < .001, \eta^2_G = .03$. A significant main effect of change size was also revealed, however sphericity was violated (W

= 0.49, p = .001), $F(3, 87) = 43.79$, p = .001, η^2G = .03. A significant but small interaction between representation and change size was also revealed, $F(3, 87) = 5.58$, p = .002, , η^2_G = .002. Polynomial contrasts revealed a significant linear trend (p = .02) over change size, with participants' confidence ratings increasing over change sizes for the averages and arrays. That is, confidence also increased as change size increased.

CHAPTER 3

Experiment 2: Image Alteration

Experiment 2 asks the same question as Experiment 1: *Did you see a change?* But this time, instead of fixing the number of images and manipulating the number of changes, I investigate sensitivity to change by altering the number of images displayed. So I still manipulate change size, but by fixing the number of changes to 1 and manipulated the number of images in the display. Therefore, either two, four, eight or sixteen images are shown in any array or average trial but on the change trials, only 1 image (selected at random) changes on the second presentation.

3.1 Method

3.1.1 Participants. The same group of participants completed Experiment 1 and 2 concurrently, with some opting to take a short few-minute break between the two experiments. To randomise the order in which participants' completed each experiment they reached into an opaque bag to randomly select one of two pods labelled either "1" or "2"to determine which experiment they would complete first.

3.1.2 Procedure. The procedure for generating the arrays and averages, and the number of trials and procedure for presenting them to participants was identical to Experiment 1, with the exception of the variation in the change size manipulation. The instructional video for Experiment 2 can be viewed here:

https://www.youtube.com/watch?v=a9swkqqAZ8Q

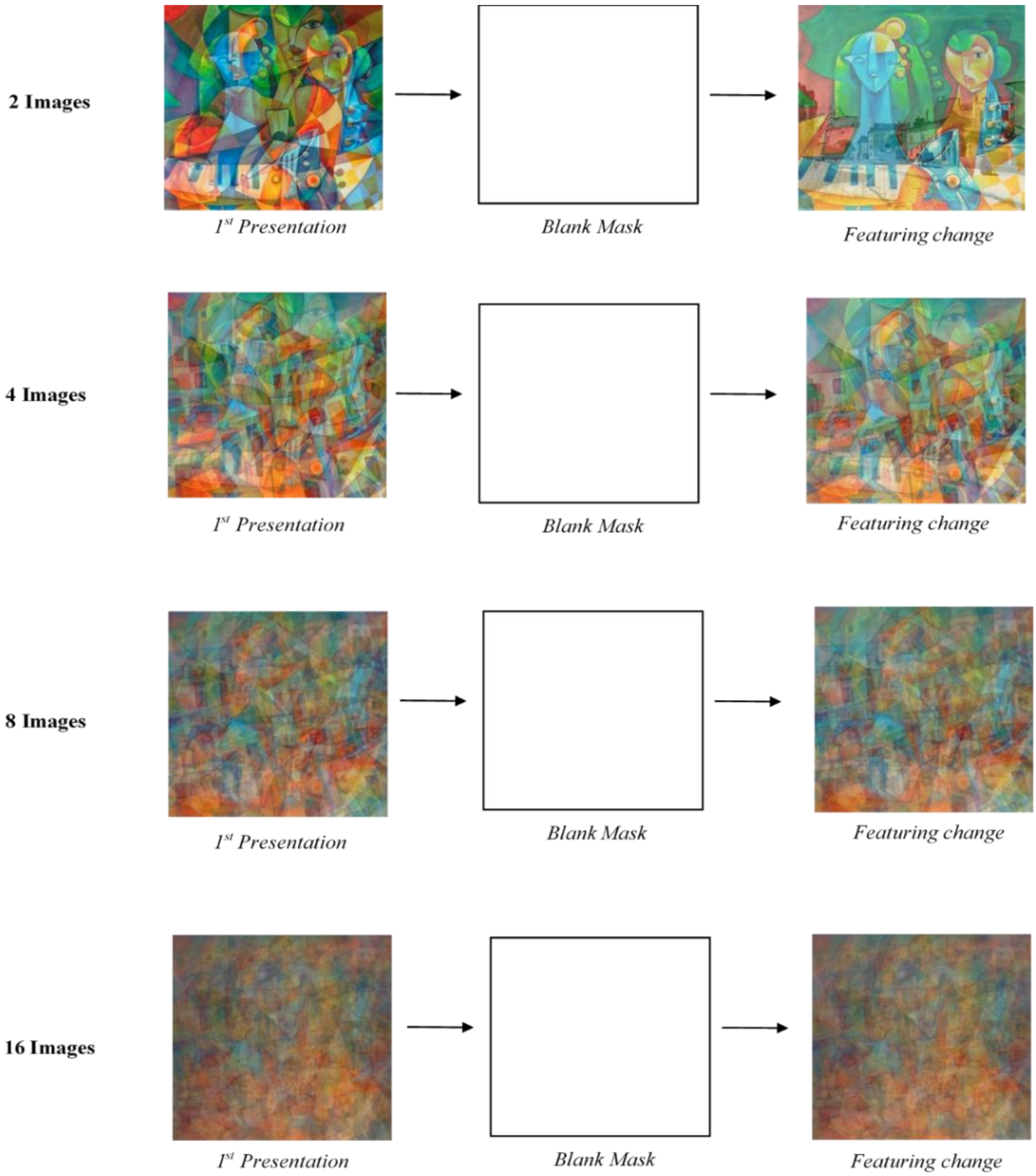


Figure 8. Number of Images in each condition of the Averages featuring Cubist paintings.

Figure 9. Number of Images in each condition of the Arrays featuring Cubist paintings.

3.2 Results

The same analyses conducted in Experiment 1, were repeated for Experiment 2 to examine the effect of representation (averages versus arrays) and change size (1 change out of 2, 4, 8, or 16 images) on participants' change detection ability and confidence. Note that for Experiment 2, change size is now operationalised by varying the number of images displayed on screen with a single changing image, rather than by varying the number of changing images in a fixed display of 16 images.

3.2.1 Descriptives. There were a few outlying observations present in the data, (see Figure 11), however these were retained as no participants' displayed a pattern of responding that met the prespecified data exclusion criteria. Unlike in Experiment 1, the assumption of sphericity was not violated for the change size manipulated. Therefore, no corrections were applied and parametric analyses were carried as planned in the preregistration.

Overall, participants performed better in the averages (M = .69, SD = .12) than the arrays (M = .61, SD = .14). Table 2 shows participant sensitivity and proportion correct in each condition. To check for floor effects, a paired t-test was conducted comparing participants' AUC scores to the simulated randomly responding participants. Participants were indeed performing significantly above chance; $t(239) = -11.28$, $p < .001$, 95% CI [-0.19, -0.12]. This result suggests that an overall floor effect was not present.

While people were better at detecting changes in the averages overall, discriminability with the 16 image condition (the condition where change size was the *smallest*; 1 image changed out of 16) produced similar sensitivity scores for both averages (M = .44, SD = .12) and arrays (M = .45, SD = .16). Additionally, as change size increased, so did participants' sensitivity to changes within the visual display. For example, the change size or proportion of change, is largest in the 2-Images condition half of the display changes (e.g., one of two

Cubist paintings swapped for an Impressionist). Whereas, change size in the 16-image condition results in a 6.25% alteration to the display, which is more difficult to observe.

Table 2

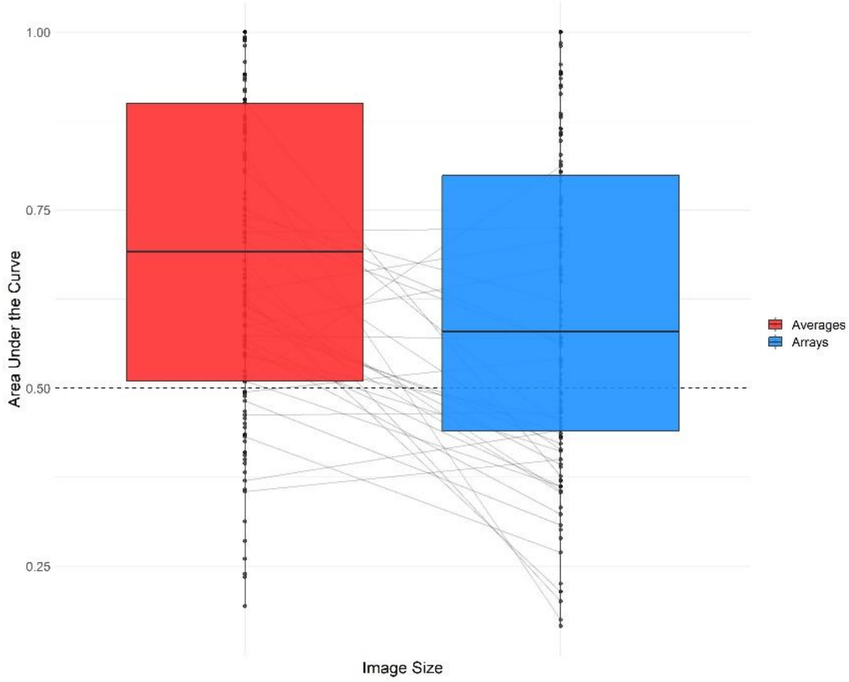
Experiment 2 Discriminability and Proportion of Correct Scores across Representation and Number of Images

	Discriminability (AUC)		Proportion Correct	
	Averages	Arrays	Averages	Arrays
Number of Images	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)
2 Images	.94 (.09)	.85 (.14)	.90 (.09)	.87 (.09)
4 Images	.79 (.13)	.67 (.16)	.79 (.09)	.73 (.10)
8 Images	.61 (.13)	.48 (.15)	.68 (.08)	.60 (.08)
16 Images	.44 (.12)	.45 (.16)	.55 (.08)	.56 (.10)

Note. The AUC scores refer to mean (and standard deviation) Area Under the Curve for each condition in experiment

3.2.2 Discriminability. There was a significant effect of representation type on participants' sensitivity to change, $F(1, 29) = 17.05$, $p < .001$, $\eta^2_G = .09$, with a medium to large effect size (see Figure 10A). Additionally, there was a significant effect of image size on participants' sensitivity to change, $F(3, 87) = 273.16$, $p < .001$, $\eta^2_G < .63$, featuring a large effect size (see Figure 10B).

A.



B.

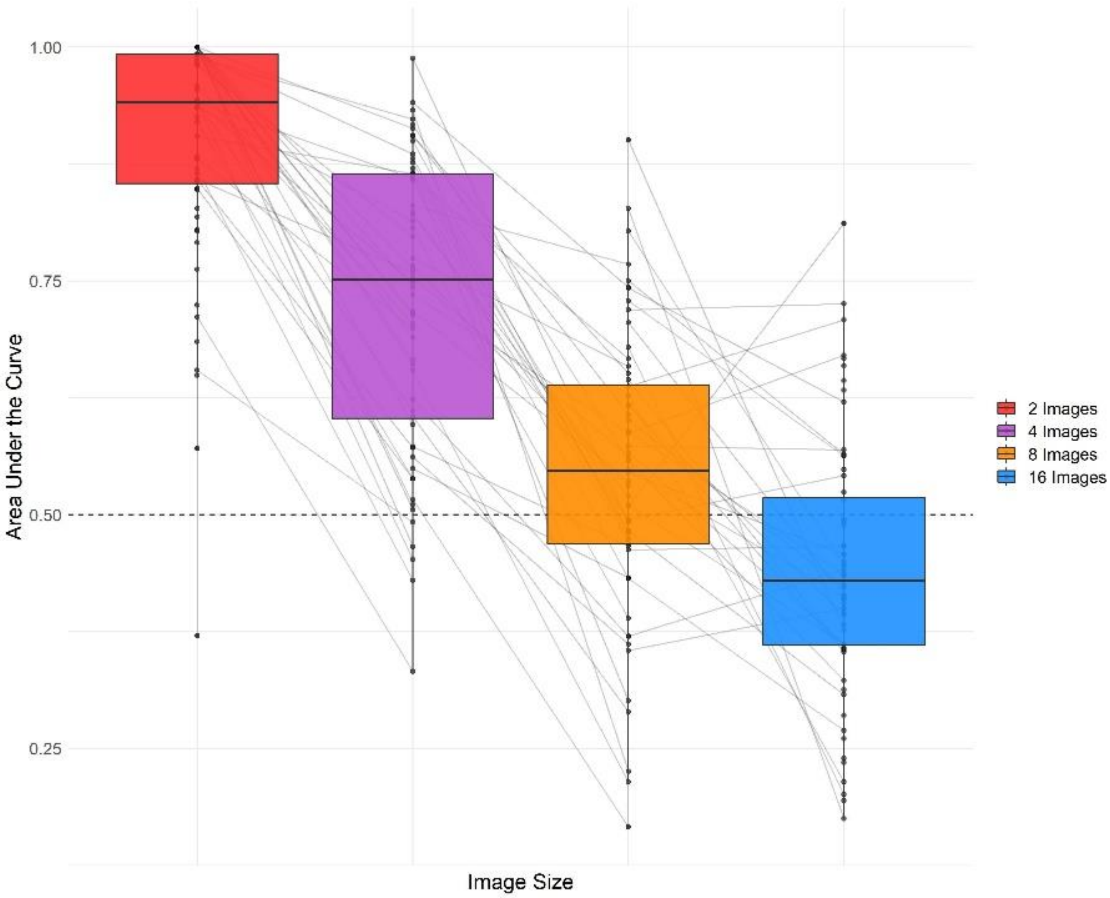


Figure 10. Panel depicts (A) main effect of participants' discriminability in performance across both representations, collapsed across number of images and domain (Scenes and Paintings); and (B) main effect of discriminability in sensitivity to the number of images, collapsed across representation (Arrays and Averages) and domain (Scenes and Paintings).

Moreover, a significant interaction between representation and change size was found with a medium effect size, $F(3, 87) = 6.17$, $p < .001$, $\eta^2_G = .04$ (see Figure 11). Polynomial contrasts further revealed a significant linear ($p < .001$) and quadratic trend ($p = .046$) over change sizes with the arrays, revealing participants' sensitivity to detecting change across image size conditions decreased exponentially as the number of images increased. That is, greater sensitivity was observed with lower number of images in the display (larger change size), with an exponential decrease in performance as the number of images increased (smaller change size). A significant linear trend ($p < .001$) was also observed with the averages, with participants' sensitivity in the 2-image condition resembling near perfect performance ($M = .94$, $SD = .09$).

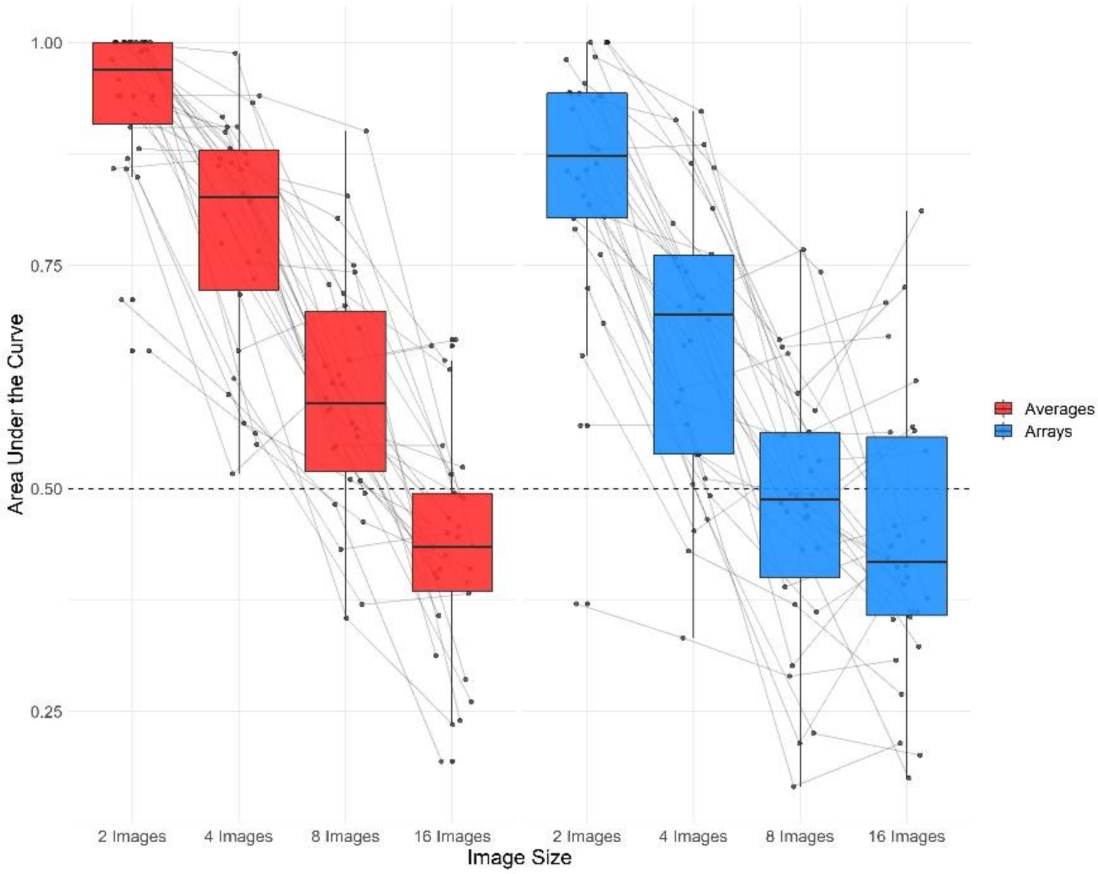


Figure 11. Interaction effect between Representation and Number of Images for Experiment 2.

3.2.3 Confidence. Overall, participants were more confident at detecting changes in the averages (*M* = 3.54, *SD* = 1.31) compared to the arrays (*M* = 3.29, *SD* = 1.36). However, the 2-Image condition in the arrays produced the highest confidence (*M* = 4.43, *SD* = 1.32), as compared to the 2-Images condition in the averages (*M* = 4.27, *SD* = 1.15), and all other image size conditions (see Figure 12). Confidence at detecting changes in the averages comprising 16 images (*M* = 3.01, *SD* = 1.39) was higher than both the 8-Image (*M* = 2.81, *SD* = 1.38) and 16-Image (*M* = 2.52, *SD* = 1.28) conditions within the arrays.

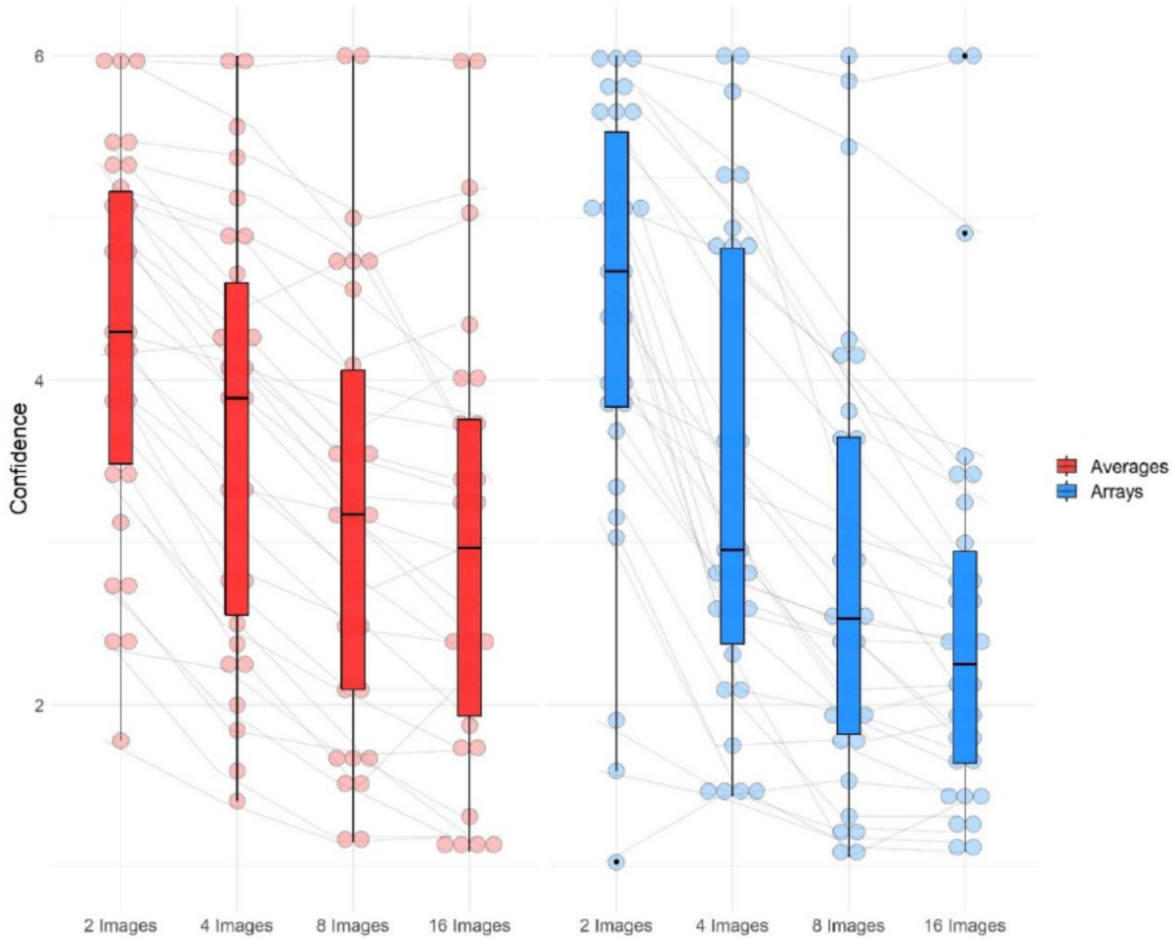


Figure 12. Confidence across image size in Experiment 2.

A within-subjects ANOVA revealed a small but significant main effect of representation on confidence, $F(1, 29) = 4.83$, $p = .04$, $\eta^2_G = .01$. A significant main effect of image size was also revealed; $F(3, 87) = 105.05$, $p < .001$, $\eta^2_G = .18$, featuring a large effect. A significant but small interaction between representation and change size was also uncovered, $F(3, 87) = 9.52$, $p < .001$, $\eta^2_G = .01$. Confidence increased as change size (proportion of display changed) also increased, thus highest when less images were presented on-screen. Polynomial contrasts revealed a significant linear trend ($p < .001$) over display size, with participants' confidence ratings decreasing over change sizes for the averages and arrays. That is, confidence decreased as display size increased.

Notably, whilst participants reported being most confident in the arrays 2-image condition, they were more accurate in discriminating change and obtaining proportion correct within the averages 2-image condition than arrays.

CHAPTER 4
Discussion

My thesis brings together gist perception, ensemble coding, and VSTM research by exploring human ability to detect global and local changes via a one-shot detection task. The research seeks to provide further understanding of visual perception and offer insight into the nature of VSTM capacity. Question remains in the literature as to how rich or sparse our perception is. Subjective experience appears rich, however past literature identifies limits, and even blindness to large scene changes (Cohen, Ostrand & Frontero, 2019; Miller, 1994). To answer this, I explored human sensitivity to detecting changes in scenes and paintings by implementing summarised scenes and series of images. We varied change size to better isolate the conditions under which change detection may be better or worst in the averages and arrays to identify the sweet spot of change detection.

One view of ensemble coding contends that observers represent their world as an ensemble statistic (Cohen et al., 2016a). While VSTM research contends there are only three to four items held in memory at a time (Cowan, 2000). Brady and Tenenbaum (2012) argue people encode both, overall gist of a scene (ensemble statistic) and detailed information about objects. Their 2012 research provided evidence that visual memory capacity can retain approximately 24 items. When their 5×5 arrays included more structure, visual memory capacity increased from the standard estimates of three-four slots.

These accounts of VSTM from change detection studies suggest summary statistics *and* individual items are held in memory, but estimates of capacity tend to differ across studies. As this project was exploratory, change size was manipulated in to gauge how different representation effects detection of changes with varying magnitude. This was investigated with two twin experiments: by shifting the *number of changes* within a display, and *number of images*.

4.1 Findings

Prediction 2 proposed participants will be more sensitive to changes in the arrays than averages. Contrary to this, in both perceptual experiments participants were more sensitive to detecting changes in the averages than the arrays. Despite working memory constraints, ensemble coding gives ruse to the feeling of richness and a possible explanation to bridging the gap between claims that sparse or detailed views are at play. These findings lend support to research that affirms people perceive far less than intuitively believed, (Cohen et al., 2016b; Haberman & Whitney, 2012; Huan et al., 2017). Therefore, redundant features, such as every individual leaf that makes up a tree, is seen past and a statistical summary is created. This is based upon various ensemble features including orientation, density, position and size (Jackson-Nielsen et al., 2017; Geisler, 2008). This offers a possible explanation into why the averages condition resulted in better discriminability performance. People see the gist and are slower to see the detail, thus extra detail in the arrays does not suggest adding much during brief periods of exposure as seen in the experiment.

People have a remarkable ability to often accurately capture the gist of a scene, capturing the statistical visual summary almost instantaneously and frequently between 100 to 200ms (Oliva & Torralba, 2006; Whitney & Leib, 2018). The nature of the averages means the images, when presented for 150ms, are already summaries of their subsequent individual scenes from the arrays. The averages combined with brief exposure durations may have made the ensemble even more salient and thus easier to encode. By averages having computed the mean global image features to create a statistical visual summary, this may prove advantageous in providing a cognitive shortcut in determining the scene's average.

Additionally, visual perception literature highlights low-level global features, such as hue and spatial position, play a role in the formation of gist (Oliva & Torralba, 2006). It is often stated that global features are perceived first in a scene, believed to guide which local details are attended to (Ericson, Beck & Lamsweerde, 2016; Navon, 1977). Participants in this project were presented with displays for less than half a second. Therefore, there purposely was not enough time to make an eye movement or blink, compelling the visual system to capture the gist - the scene at a glance, and straining the ability to capture many local details or encode beyond short-term visual memory. Given the averages are a condensed synopsis of the arrays, these low-level ensemble features may provide an advantage in visual memory encoding. In the current project, images were artificially summarised lending cognitive aid to obtaining a gist, suggesting an explanation into the results achieved.

Experiment 1 explored observers' sensitivity to change size by the varying number of images that change in display of sixteen paintings of scenes. Significant linear trends in discriminability were observed in for both averages and arrays, revealing that the increment performance increase remains the same across each of the change levels. As the proportion of change increased, so did discriminability. While the interaction between representation type (Averages, Arrays) and change size (1, 2, 4, 8 images changed out of 16) reported small effect size, notably there was a very large effect size for number of changes within the display upon discriminability. This supports prediction 1 which proposed participants will be more sensitive to change as the size of change increases. This result suggests that variance in the ensemble is useful when identifying statistical visual outliers.

Ensemble variance is extracted for low-level (e.g., orientation, Norman, Heywood & Kentridge, 2015) and mid-level ensemble features (e.g., size, Solomon, Morgan & Chubb, 2011). This can lead to stark differences in the summary between two visual presentations, with observers having been found to accurately discriminate and reproduce a statistical moments (Whitney & Yamanashi-Lieb, 2018). The increasing sensitivity to changes with increasing change size suggest that more variance provides a larger disruption to the original ensemble participants may have encoded. This result lends support to the fact participants could accurately discriminate between change and no-change trials in the larger change size conditions, relatively well above chance. These results suggest creating an ensemble, or statistical summary assists VSTM capacity.

Participants confidence ratings also followed a similar pattern to their discriminability in the two experiment. Even though people often believe detecting changes in scenes will be easy, they are often surprised to find it is not (Levin, 2000). Confidence increased as the proportion of change increased, supporting prediction 3. When the change became more prominent and harsher in the visual display, participants indicated it was ‘more obvious’. Confidence ratings mirrored discriminability performance overall, therefore suggesting observers are consistent in gauging their relative ability across the different conditions. This finding is consistent with Brady & Tenenbaum’s (2012) arrays task, whereby observers were also reliable in stating which changes they found difficult or easy to detect.

Experiment 2 further explored observers VSTM capacity for complex images through change detection by manipulating the number of images presented in the display. Should people correctly discriminate between a change or no-change trial, it suggests they are holding a reasonably accurate gist in mind. Therefore, may assist in determining a decision regarding the second trial presentation. Experiment 2 found participants could reliably discriminate whether a change occurred or not (above chance) at 2-image and 4-image display sizes. While the 8-images were above chance, participants performed less reliably in this image size condition within the arrays. Participants were more sensitive to changes as the proportion of change increased; that is, as the number of images in the display increased with a single changing image. Likewise, participants were more confident as the size of the change increased.

Notably, participants were more confident in the arrays 2-image, even though their discriminability and proportion correct with the averages was higher. This finding suggests that more detail is not necessarily better in detecting changes at a glance. This seems counterintuitive given, when trying to recognise alterations, one would want as much information as possible (Jackson-Neilsen et al., 2017). However, in both experiments I observed better discriminability within the averages representation than arrays, contrary to my second predition that proposed participants will be more sensitive to changes in the arrays than the averages. It appears people are more sensitive to detecting changes that disrupt the holistic or global representation of the display, compared with those that only disrupt local features and objects.

In real-world scenes, people not only encode information about specific objects but semantic gist of a scene too. The findings from both experiments support previous research that has modelled scene gist and found local, item-based information not necessary to determine gist (Koehler & Eckstein, 2017; Oliva & Torralba, 2006). The averages showcase various ensemble features of texture, hue and orientation, suggesting these features are possibly processed more easily given the limited time to encode into memory.

Moreover, results from this current project supports of VSTM that suggest people encodes global visual information more robustly than local detail-based features (Vidal, Gauchou, Tallon-Baudry, & O'Regan, 2005; Luck & Vogel, 1997). Changes in the arrays may have gone unnoticed more frequently than the averages as the local internal change may not have been a large enough disruption to pique observer sensitivity.

Interestingly, both experiments captured one change within 16 images (Experiment 1: arrays, 1-change; Experiment 2: arrays, 16-images) with mean accuracy falling just below chance for both. Moreover, feature-based theory may suggest an explanation in the quadratic trend revealed within the arrays condition of Experiment 2. The steep decline in performance sensitivity from 2-images to 4-images and 8-images perhaps illuminates the limits of VSTM capacity. Change detection tasks impose processing demands on visual memory, beyond those of object recognition with even dramatic changes going unnoticed (Gaspar, Neider, Simons, McCarley & Kramer, 2013). However, the ability to encode the pre-change display and then subsequently decide if it matches to a second display trial is one that is remarkable. Yet is is not without limits as efficient cognitive processes can lead to missing large, often dramatic changes to the visual environment.

4.2 Strengths and Limitations

Several strengths of the research stand out as constructive to the current literature. My current research contributes empirical evidence to the emerging field by further understanding ways the visual system exploits redundancy in real-world scenes to represent a large amount of information presented briefly. I test people’s ability to retain global versus local information in VSTM by examining their ability to detect changes in visual averages (sum total divided by number of items) and through their sensitivity, capture the various capacities of VSTM.

I expect the findings to generalise to other natural image sets, such as amusement parks, classrooms and golf courses, as used by Konkle, Brady, Alvarez & Oliva (2010). I used paintings, urban and natural scenes for generality given real-world visual experiences often are complex and involve scenes from such environments.

Moreover, a power analysis to determine sample size indicated high chance of detecting an effect with 30 observers. These participants were a sub sample from the general public and therefore, it is believed the findings can also be generalised to the broader Australia population too.

The project does have its limitations. Firstly, image size and transparency were not matched between arrays and averages. A trade-off was made between matching the conditions on display size. Ultimately it was decided to match according to pixels with the overall display equating to 800×800 pixels. This was to ensure the stimuli were similarly distributed across the visual field in each condition and control for display. If representations were matched to image size and averages set to the individual array size of 200×200 pixels, the average display would be markedly smaller than the array display.

Secondly, to have captured a more direct comparison of discriminability between visual representation of a summary or series of images, it would have been beneficial to randomise images within the arrays for the second presentation. By scrambling the location of each individual array item, this would create a disruption to the global representation and structure of the display.

4.3 Future investigation

There are a number of future lines of research that follow from this project. Firstly, future studies may attempt to scramble the location of images presented in arrays to investigate what occurs when global representation is further disturbed. People may still have been detecting changes in the arrays as the global structure of the array was left intact. Despite arrays providing a degree of local and global information, people were still better at detecting changes in the averages.

Secondly, Cohen (2019) argues participants may not be aware of a target's precise location, but one may still hold some imprecise knowledge regarding its location. From this, it may be useful to capture qualitative claims from observers who may not precisely be aware of a change, although voice their thinking process to capture approximate awareness of items (e.g- change located top right, or hue, colour or texture seems to have been altered). Alternatively, if observers can provide multiple responses instead of a single forced-choice response, this may shed more light on what participants consciously remember.

4.4 Implications and Conclusion

By employing differing visual representations to explore gist and by manipulating change size; I demonstrated that summary representation (averages) are sufficient and more detail (arrays) are not necessary for accurate detection of changes in natural scenes and paintings. In fact, the less detailed average representations produced significantly better change detection performance.

This result is counterintuitive given that participants had more information at their disposal in the arrays compared to the averages to draw connections with across the presentations. Additionally, these results provide empirical evidence to better understanding the capacities of perceptual experience and contributes to visual perception theory. As McClelland and Bayne (2016) propose, sparsity still encapsulate key information than can be used to inform ensemble judgement. VSTM research grows to understand the underpinnings of our cognitive mechanisms. This project demonstrated that people seem to be better able to retain summary representations of a display, which provides some insight into how people represent their visual environment in VSTM.

Change detection is important and used in much of our everyday life. This includes deciding when to safely merge into traffic by judging speed, location and direction of other cars, or spotting a child that has run onto the road to retrieve their ball (Rensink, 2002). When significant changes in scenes or objects go undetected, this phenomenon of change blindness can produce inherently consequential results. For example, air traffic controllers simultaneously juggle visual searches, situational evaluations and communicate with relevant personnel while organising the flow of air traffic (Imbert et al., 2014). Should change blindness occur and a significant piece of visual information be missed, this can lead to fatal accidents.

Further exploring how we see the world around us, aids theoretical understanding of how people remember their visual experiences and can benefit applied settings too. An understanding in how our visual system operates, alongside the potential cognitive shortcuts that may lead to pitfalls in judgement or accuracy, can be used to better understand and assist applied settings. For example, change detection may impact eyewitness testimony statements (Simons, & Ambinder, 2005). Change blindness occurs in a significant proportion of observers when a mask or noise has interrupted view of a scene, such as a car blocking the view of a suspected perpetrator or person of interest. This may lead observers to reporting inaccurate details, such as the characteristics of a person and risks causing a potential miscarriage of justice (Lindsay, Wells & Rumpel, 1981). Insight into how our visual system works can enable critical thinking when such real-world situations involve careful and hold significant consequences should an error in judgement be made.

Ultimately, while we may believe we are seeing an incredibly rich, detailed world, these findings suggest that this richness is underpinned by surprisingly sparse global representations. For more detail is not necessarily better, nor more advantageous in detecting changes within complex natural scenes, or in generating a gist of a visual ensemble. It appears our sparse perception may not be so sparse after all.

Contributions

The project idea was jointly conceived by my supervisor and I, whereby we worked together to refine the concept and design of both experiments. I carried out the literature review and assembled the ethics application for approval from the School of Psychology. I preregistered the project on the Open Science Framework with guidance from my supervisor.

My supervisor programmed the experiments and I piloted them seven times to test the user interface along with various averaging methods. I contributed to the generation of the average images, the preparation and recording of the video instructions. I also ran the 60 computer simulations across both experiments.

I coordinated and carried out all of the data collection for both experiments. I made a new Google email address for the experiment to assist scheduling of participant sessions. Additionally, I made a spreadsheet to track the sequence number for each trial and to verify participant compensation had been received. I created participant recruitment posters and distributed them across social media and North Terrace campus. I also recruited participants via the RPS. My supervisor programmed an application for extracting the raw data from the 60 individual .txt files, and provided guidance on using R and R Markdown. I performed all data analyses in R (version 3.5.2) using RStudio (version 1.2.1335).

References

Alvarez, G. A., & Oliva, A. (2008). The Representation of Simple Ensemble Visual Features Outside the Focus of Attention. *Psychological Science, 19*(4), 392-398

Alvarez, G. A., & Oliva, A. (2009). Spatial ensemble statistics are efficient codes that can be represented with reduced attention. *Proceedings of the National Academy of Sciences, 106*(18), 7345-7350.

Bakeman, R. (2005). Recommended effect size statistics for repeated measures designs. *Behavior Research Methods, 37*(3), 379-384.

Bateman, J. E., Ngiam, W. X. Q., & Birney, D. P. (2018). Relational encoding of objects in working memory: Changes detection performance is better for violations in group relations. *PLoS ONE, 13*(9), e0203848.

Brady, T. F., & Alvarez, G. A. (2011). Hierarchical encoding in visual working memory: ensemble statistics bias memory for individual items. *Psychological Science, 22*(3), 384-392.

Brady, T. F., & Tenenbaum, J. B. (2013). A Probabilistic Model of Visual Working Memory: Incorporating Higher Order Regularities Into Working Memory Capacity Estimates. *Psychological Review, 120*(1), 85-109.

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

Brady, T. F., Konkle, T., Oliva, A., & Alvarez, G. A. (2009). Detecting changes in real-world objects: The relationship between visual long-term memory and change blindness. *Communicative & Integrative Biology, 2*(1), 1-3.

Brennan, P., Gandomkar, Z., Ekpo, E., Tapia, K., Trieu, P., Lewis, S., . . . Evans, K. (2018). Radiologists can detect the 'gist' of breast cancer before any overt signs of cancer appear. *Scientific Reports, 8*(1), 8717.

Chabris, C. F., Weinberger, A., Fontaine, M., & Simons, D. J. (2011). You do not talk about Fight Club if you do not notice Fight Club: Inattentional blindness for a simulated real-world assault. *i-Perception*, *2*(2), 150–153.

Cohen, M. A., & Chun, M. M. (2017). Studying consciousness through inattentional blindness, change blindness, and the attentional blink. In A. S. Schneider & M. Velmans (Eds.), *The Blackwell Companion to Consciousness* (2nd ed.) (pp. 539-547). Retrieved from http://www.michaelacohen.net/uploads/5/9/0/7/59073133/ch38.pdf

Cohen, M. A., Dennett, D. C., & Kanwisher, N. (2016a). Ensemble Perception, Summary Statistics, and Perceptual Awareness: A Response. *Trends in Cognitive Sciences, 20*(9), 643-644.

Cohen, M. A., Dennett, D. C., & Kanwisher, N. (2016b). What is the Bandwidth of Perceptual Experience? *Trends in Cognitive Sciences*, *20*(5), 324-335.

Cohen, M.A. (2019) What is the true capacity of visual cognition? *Trends in Cognitive Sciences*. 23, 83-86.

Cohen, M., Ostrand, C., & Frontero, N. (2019). Characterizing a snapshot of perceptual experience. *Journal of Vision*, *19*, 2-22.

Cowan, N. (2001). The Magical Number 4 in Short-Term Memory: A Reconsideration of Mental Storage Capacity. *The Behavioral and Brain Sciences*, *24*, 87-114.

Ericson, J., Beck, M., & Lamsweerde, M. (2016). Binding global and local object features in visual working memory. *Attention, Perception, & Psychophysics, 78*(1), 94-106.

Field, A., Miles, Jeremy, & Field, Zoë. (2012). *Discovering statistics using R*. Los Angeles: SAGE.

Franconeri, S. L., Alvarez, G. A., & Enns, J. T. (2007). How many locations can be selected at once? *Journal of Experimental Psychology: Human Perception and Performance, 33*(5), 1003-1012.

Gaspar, J. G., Neider, M. B., Simons, D. J., McCarley, J. S. & Kramer, A. F. (2013). Change detection: training and transfer. *PLoS ONE, 8*(6), E67781.

Geisler, W. (2008). Visual Perception and the Statistical Properties of Natural Scenes. *59*(1), 167-192.

Haberman, J., & Whitney, D. (2009). Seeing the Mean: Ensemble Coding for Sets of Faces. *Journal of Experimental Psychology: Human Perception and Performance, 35*(3), 718-734.

Hanley, J. A., & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology, 143*(1), 29-36.

Haun, A. M., Tononi, G., Koch, C., & Tsuchiya, N. (2017). Are we underestimating the richness of visual experience? *Neuroscience of Consciousness, 2017*(1), 1-4.

Howe, P. D. L. (2017). Natural scenes can be identified as rapidly as individual features. *Attention, Perception, & Psychophysics, 79*(6), 1674-1681.

Howe, P. D. L., Cohen, M. A., Pinto, Y., & Horowitz, T. S. (2010). Distinguishing between parallel and serial accounts of multiple object tracking. *Journal of Vision*, *10*(8), 11.

Hyman Jr, I. E., Boss, S. M., Wise, B. M., McKenzie, K. E., & Caggiano, J. M. (2010). Did you see the unicycling clown? Inattentional blindness while walking and talking on a cell phone. *Applied Cognitive Psychology, 24*(5), 597-607.

Imbert, J., Hodgetts, H., Parise, R., Vachon, F., Dehais, F., & Tremblay, S. (2014). Attentional costs and failures in air traffic control notifications. *Ergonomics, 57*(12), 1817-1832.

Jackson-Nielsen, M., Cohen, & Pitts. (2017). Perception of ensemble statistics requires attention. *Consciousness and Cognition, 48*, 149-160.

Jiang, Y., Olson, I. R., & Chun, M. M. (2000). Organization of Visual Short-Term Memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*(3), 683-702.

Koehler, K., & Eckstein, M. P. (2017). Beyond scene gist: Objects guide search more than scene background. *Journal of Experimental Psychology: Human Perception and Performance, 43*(6), 1177-1193.

Konkle, T., Brady, T. F., Alvarez, G. A., & Oliva, A. (2010). Scene Memory Is More Detailed Than You Think: The Role of Categories in Visual Long-Term Memory. *Psychological Science, 21*(11), 1551-1556.

Levin, D., Momen, N., Drivdahl, S., & Simons, D. (2000). Change Blindness Blindness: The Metacognitive Error of Overestimating Change-detection Ability. *Visual Cognition, 7*(1-3), 397-412.

Lindsay, R. C., Wells, G. L., & Rumpel, C. M. (1981). Can people detect eyewitness-identification accuracy within and across situations? *Journal of Applied Psychology, 66*(1), 79-89.

Luck, S. J., & Vogel, E. K. (2013). Visual working memory capacity: from psychophysics and neurobiology to individual differences. *Trends in Cognitive Sciences, 17*(8), 391-400

Massoni, S., Gajdos, T., & Vergnaud, J.-C. (2014). Confidence measurement in the light of signal detection theory. *Frontiers in Psychology, 5*(1455), 189-198.

Maurer, D., Le Grand, R., & Mondloch, C. J. (2002). The many faces of configural processing. *Trends in Cognitive Sciences*, *6*(6), 255-260.

McClelland, T., & Bayne, T. (2016). Ensemble Coding and Two Conceptions of Perceptual Sparsity. *Trends in Cognitive Sciences, 20*(9), 641-642.

Miller, G. (1994). The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information. *Psychological Review, 101*(2), 343-352.

Murphy, G., & Murphy, L. (2018). Perceptual load affects change blindness in a real-world interaction. *Applied Cognitive Psychology, 32*(5), 655-660.

Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology*, *9*(3), 353-383.

Noe, A., Pessoa, L., & Thompson, E. (2000). Beyond the Grand Illusion: What Change Blindness Really Teaches Us About Vision. *Visual Cognition, 7*(1-3), 93-106.

Norman, L. J., Heywood, C. A., & Kentridge, R. W. (2015). Direct encoding of orientation variance in the visual system. *Journal of Vision*, *15*(4), 1-14.

Oliva, A., & Torralba, A. (2001). Modeling the Shape of the Scene: A Holistic Representation of the Spatial Envelope. *International Journal of Computer Vision, 42*(3), 145-175.

Oliva, A., & Torralba, A. (2006). Building the gist of a scene: the role of global image features in recognition. *Progress in Brain Research, 155*, 23-35.

Pessoa, L., & Ungerleider, L. G. (2004). Neural Correlates of Change Detection and Change Blindness in a Working Memory Task. *Cerebral Cortex*, *14*(5), 511-520

Rensink, R. (2000). Seeing, sensing, and scrutinizing. *Vision Research*, *40*(10-12), 1469-1487.

Rensink, R. A. (2002). Change detection. *Annual Review of Psychology*, *53*, 245-277.

Rensink, R. A., O'Regan, J. K., & Clark, J. J. (1997). To See or not to See: The Need for Attention to Perceive Changes in Scenes. *Psychological Science, 8*(5), 368-373.

Rousselet, G., Joubert, O., & Fabre-Thorpe, M. (2005). How long to get to the “gist” of real-world natural scenes? *Visual Cognition, 12*(6), 852-877.

Sampanes, A., Tseng, P., & Bridgeman, B. (2008). The role of gist in scene recognition. *Vision Research, 48*(21), 2275-2283.

Schwabe, K., Menzel, C., Mullin, C., Wagemans, J., & Redies, C. (2018). Gist Perception of Image Composition in Abstract Artworks. *Iperception, 9*(3), 1-25.

Schyns, P. G., & Oliva, A. (1994). From blobs to boundary edges: Evidence for time- and spatial-scale-dependent scene recognition. *Psychological Science, 5*(4), 195-200.

Searston, R. A., & Tangen, J. M. (2017). The Emergence of Perceptual Expertise with Fingerprints Over Time. *Journal of Applied Research in Memory and Cognition, 6*(4), 442-451

Searston, R. A., Thompson, M. B., Vokey, J. R., French, L. A., & Tangen, J. M. (2019). How low can you go? Detecting style in extremely low resolution images. *J Exp Psychol Hum Percept Perform, 45*(5), 573-584.

Simons, D. J., & Ambinder, M. S. (2005). Change Blindness: Theory and Consequences. *Current Directions in Psychological Science, 14*(1), 44-48.

Simons, D. J., & Levin, D. T. (1997). Change blindness. *Trends in Cognitive Sciences, 1*(7), 261-267

Solomon, J. A., Morgan, M., & Chubb, C. (2011). Efficiencies for the statistics of size discrimination. *Journal of Vision, 11*(12), 1-11.

Steven J. Luck, & Edward K. Vogel. (1997). The capacity of visual working memory for features and conjunctions. *Nature, 390*(6657), 279-281.

Suchow, J., Fougnie, W., Brady, D., & Alvarez, T. (2014). Terms of the debate on the format and structure of visual memory. *Attention, Perception, & Psychophysics*, *76*(7), 2071-2079.

Thompson, M. B., & Tangen, J. M. (2014). The Nature of Expertise in Fingerprint Matching: Experts Can Do a Lot with a Little. *PLoS ONE, 9*(12), e114759.

Torralba, A. (2009). How many pixels make an image? *Visual Neuroscience, 26*(1), 123-131.

Vidal, J. R., Gauchou, H. L., Tallon-Baudry, C., & O'Regan, J. K. (2005). Relational information in visual short-term memory: The structural gist. *Journal of Vision*, *5*(3), 244-256.

Vokey, J. R. (2016). Single-Step Simple ROC Curve Fitting via PCA. *Canadian Journal of Experimental Psychology, 70*(4), 301-305.

Whitney, D., & Yamanashi Leib, A. (2018). Ensemble Perception. *Annual Review of Psychology, 69*(1), 105-129.

Whitney, D., Haberman, J., & Sweeny, T.D. (2014). From textures to crowds: multiple levels of summary statistical perception. *The New Visual Neurosciences*. Ed. J. S. Wener and L. M. Chalupa. Cambridge: MIT Press, 2014. 695-710.

Appendix A
Recruitment Poster

*contact information tabs redacted for deidentified marking purposes

Appendix B
SONA screenshot advertisement

Study Name	Can You Spot the Difference in Artwork?
Study Type	**Standard (lab) study** This is a standard lab study. To participate, sign up, and go to the specified location at the chosen time.
Study Status	**Visible to participants :** Approved **Active study :** Appears on list of available studies
Duration	40 minutes
Credits	1 Credits
Abstract	You're invited to participate in a lab-based study looking at how accurately people can detect changes in visual images of paintings and natural scenes.
Description	Participants will be required to complete a computer-based visual detection task at the [redacted] Lab (Level 2 Hughes building, [redacted], University of Adelaide, North Terrace Campus). The task involves two experiments that go for 15-20 minutes each and allows for a break in-between. You'll be presented with cubist and impressionist paintings, as well as scenes of natural environments (beaches, mountains, sunsets) and urban environments (skyscrapers, roads, city squares). In each trial, an image will be shown very briefly (less than a second) followed by a blank screen and then a second image also shown briefly. The second image is either identical or altered to the first image. You'll then be prompted to select whether you believe a change occurred between the two and how confident you are in this selection. All research activities will be completed within the one testing session. This study aims to explore visual short-term memory, accuracy of gist perception and visual change detection. Please note: Participants signed up in RPS are not eligible for gift cards and will be granted credit in return for participation.
Eligibility Requirements	18 years or over to participate, fluent in English, normal/corrected vision (please bring along eye glasses/contact lenses if applicable)

Appendix C
Information Sheet

Participant Information

THE UNIVERSITY of ADELAIDE

PROJECT: Exploring visual short-term memory for ensemble with change detection
RESEARCHER: [redacted] School of Psychology, University of Adelaide
SUPERVISOR: [redacted], School of Psychology, University of Adelaide
LOCATION: Hughes Building, North Terrace

What is the project about?

You are invited to participate in a study of human perception and cognition. This project is about investigating people's sensitivity to changing aspects of visual images (paintings and scenes in this experiment). We aim to provide insight into which elements of an image people are better at remembering and detecting changes to, thus exploring visual short-term working memory for complex images.

Who is undertaking the project?

This project is being conducted by student researcher, [redacted]. This research will form the basis for the degree of Bachelor of Psychological Science (Honours) at the University of Adelaide under the supervision of [redacted]. A responsible University of Adelaide staff member will be available nearby during the session. After you have completed the experiment, the researcher will discuss the study with you and explain the methodology of the experiment, the variables of interest, and will answer any questions you have.

What am I being invited to do?

You are invited to make judgements about images presented on a computer screen. You will be presented with a series of images, Cubist and Impressionist paintings, or Natural and Urban scenes. Your response time in milliseconds will be recorded on each trial, along with your judgment as to whether a change has occurred, and level of confidence in that judgment. You will also view an instructional video regarding the task prior to commencing and this will include several examples.

How much time will my involvement in the project take?

This process will take approximately 40 minutes and will take place within the School of Psychology at [redacted] Hughes Building, University of Adelaide, North Terrace Campus). The task will be accomplished within a single testing session.

Are there any risks associated with participating in this project?

There are no foreseeable risks in this study.

Can I withdraw from the project?

Your participation is completely voluntary. You are free to withdraw from the study at any time and will not be penalised in any way. If, for any reason, you do not want to continue with the experiment, simply let the researcher know. In this event you will still receive $20 in the form of a Coles/Myer gift card.

What will happen to my information?

Any information that is obtained from this study will remain entirely confidential and will be kept on a password protected computer with multiple redundant backups. The data from this experiment will be identified by a random number upon completion. You will not be identified by this random number, so your performance in this experiment will be recorded, but not associated with you personally. We plan to discuss the results at academic conferences both here and overseas, publish the data in international scientific journals, and store the data in an online open access repository, such as the Open Science Framework, for future studies and so that other researchers can easily reproduce our work. In any publication, presentation or online record, you cannot be identified. Your information will only be used as described in this participant information sheet and it will only be disclosed according to the consent provided, except as required by law.

.

Who can I contact if I have questions about the project?

If you have any questions about the project, you can contact the researchers via their university email addresses listed below:

[redacted]

What if I have a complaint or any concerns?

The study has been approved by the School of Psychology Human Research Ethics Subcommittee at the University of Adelaide (approval number H-2019-73). This research will be conducted according to the NHMRC National Statement on Ethical Conduct in Human Research (2007). For any questions about the ethical conduct of the research, please contact Dr Diana Dorstyn, Deputy Convenor of the Human Research Ethics Subcommittee in the School of Psychology (diana.dorstyn@adelaide.edu.au, phone: [redacted]. Any complaint or concern will be treated in confidence and fully investigated. You will be informed of the outcome.

Appendix D
Consent form

THE UNIVERSITY of ADELAIDE

Human Research Ethics Committee (HREC)

CONSENT FORM

1. I have read the attached Information Sheet and agree to take part in the following research project:

Title:	Exploring visual short-term memory for ensemble with change detection
Ethics Approval	[redacted]

2. I have had the project, so far as it affects me, and the potential risks and burdens fully explained to my satisfaction by the research worker. I have had the opportunity to ask any questions I may have about the project and my participation. My consent is given freely.

3. Although I understand the purpose of the research project, it has also been explained that my involvement may not be of any benefit to me.

4. I agree to participate in the activities outlined in the participant information sheet.

5. I understand that I am free to withdraw from the project at any time and that this will not affect my study at the University, now or in the future.

6. I have been informed that the information gained in the project may be published in a book, journal article, thesis, news article, conference presentations, website, online open access repository and/or report.

7. I have been informed that in the published materials I will not be identified and my personal results will not be divulged.

8. I agree to my information being used for future research purposes as follows:
   - Research undertaken by these same researcher(s) Yes ☐ No ☐
   - Related research undertaken by any researcher(s) Yes ☐ No ☐
   - Any research undertaken by any researcher(s) Yes ☐ No ☐

9. I understand my information will only be disclosed according to the consent provided, except where disclosure is required by law.

10. I am aware that I should keep a copy of this Consent Form, when completed, and the attached Information Sheet.

Participant to complete:

Name: ______________________ Signature: ________________________ Date:___________

Researcher/Witness to complete:

I have described the nature of the research to ____

(print name of participant)

and in my opinion she/he understood the explanation.

Signature: __________________ Position:__________________________ Date:___________

Appendix E
R code and output for Experiment 1: Descriptives

Area Under the Curve:

```
sum_data_exp1_p %>%
  group_by(Representation, N.Changes) %>%
  summarise(mean = mean(AUC),
            variance = var(AUC),
            SD = sd(AUC))
```

Representation <fctr>	N.Changes <ord>	mean <dbl>	variance <dbl>	SD <dbl>
averages	1 change	0.4715104	0.01873081	0.1368605
averages	2 changes	0.5819661	0.02468530	0.1571156
averages	4 changes	0.6810026	0.02914399	0.1707161
averages	8 changes	0.8417578	0.02131186	0.1459858
arrays	1 change	0.4857227	0.02190854	0.1480153
arrays	2 changes	0.4992383	0.03017296	0.1737037
arrays	4 changes	0.6140299	0.04188664	0.2046623
arrays	8 changes	0.7437435	0.03669762	0.1915662

8 rows

Proportion Correct:

```
sum_data_exp1_p %>%
  group_by(Representation, N.Changes) %>%
  summarise(
    mean = mean(mean_PC),
    variance = var(mean_PC),
    SD = sd(mean_PC)
  )
```

Representation <fctr>	N.Changes <ord>	mean <dbl>	variance <dbl>	SD <dbl>
averages	1 change	0.5666667	0.005706717	0.07554282
averages	2 changes	0.6489583	0.008329966	0.09126865
averages	4 changes	0.7208333	0.009020295	0.09497523
averages	8 changes	0.8093750	0.009250404	0.09617902
arrays	1 change	0.6020833	0.009087644	0.09532913
arrays	2 changes	0.6125000	0.012769397	0.11300175
arrays	4 changes	0.6781250	0.014692214	0.12121145
arrays	8 changes	0.7604167	0.015445402	0.12427953

8 rows

Confidence:

```
sum_data_exp1_p %>%
  group_by(Representation, N.Changes) %>%
  summarise(
    mean = mean(mean_conf),
    variance = var(mean_conf),
    SD = sd(mean_conf)
  )
```

Representation <fctr>	N.Changes <ord>	mean <dbl>	variance <dbl>	SD <dbl>
averages	1 change	3.040625	2.137537	1.462032
averages	2 changes	3.136458	2.031283	1.425231
averages	4 changes	3.360417	2.195699	1.481789
averages	8 changes	3.768750	1.914170	1.383535
arrays	1 change	2.743750	1.748545	1.322326
arrays	2 changes	2.708333	1.612024	1.269655
arrays	4 changes	2.869792	1.614095	1.270470
arrays	8 changes	3.133333	1.727838	1.314472

8 rows

Appendix F
R output for Experiment 1: Discriminability Analyses

t-test

```
t.test(
  sum_data_exp1$AUC[sum_data_exp1$Mode == "sim"],
  sum_data_exp1$AUC[sum_data_exp1$Mode == "participant"],
  paired = TRUE,
  alternative = "two.sided"
)
```

```
	Paired t-test

data:  sum_data_exp1$AUC[sum_data_exp1$Mode == "sim"] and sum_data_exp1$AUC[sum_data_exp1$Mode == "participant"]
t = -9.3974, df = 239, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.1792763 -0.1171397
sample estimates:
mean of the differences
              -0.148208
```

```
t.test(
  sum_data_exp1$mean_PC[sum_data_exp1$Mode == "sim"],
  sum_data_exp1$mean_PC[sum_data_exp1$Mode == "participant"],
  paired = TRUE,
  alternative = "two.sided"
)
```

```
	Paired t-test

data:  sum_data_exp1$mean_PC[sum_data_exp1$Mode == "sim"] and sum_data_exp1$mean_PC[sum_data_exp1$Mode == "participant"]
t = -16.206, df = 239, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.1882403 -0.1474367
sample estimates:
mean of the differences
             -0.1678385
```

ANOVA and Mauchly's Test for Sphericity

```
options(contrasts=c("contr.sum","contr.poly"))
ezANOVA(data=sum_data_exp1_p, dv=AUC, wid= Participant, within= .(Representation, N.Changes))
```

```
$ANOVA
```

	Effect <chr>	DFn <dbl>	DFd <dbl>	F <dbl>	p <dbl>	p<.05 <chr>	ges <dbl>
2	Representation	1	29	6.484217	1.645438e-02	*	0.03044391
3	N.Changes	3	87	108.999978	2.264942e-29	*	0.34351487
4	Representation:N.Changes	3	87	4.348724	6.651841e-03	*	0.01699940

3 rows

```
$`Mauchly's Test for Sphericity`
```

	Effect <chr>	W <dbl>	p <dbl>	p<.05 <chr>
3	N.Changes	0.6487593	0.03497351	*
4	Representation:N.Changes	0.8626141	0.53578424	

2 rows

```
$`Sphericity Corrections`
```

	Effect <chr>	GGe <dbl>	p[GG] <dbl>	p[GG]<.05 <chr>	HFe <dbl>	p[HF] <dbl>	p[HF]<.05 <chr>
3	N.Changes	0.7602505	6.669828e-23	*	0.8286479	9.490893e-25	*
4	Representation:N.Changes	0.9144634	8.496837e-03	*	1.0194475	6.651841e-03	*

2 rows

Polynomial contrasts

Averages

```
trend_averages <- sum_data_exp1  %>%
  filter(Representation == "averages")

contrasts<-contr.poly(4,c(1,2,4,8))
summary(lm(AUC ~ N.Changes, data=trend_averages, contrasts = list(f = contrasts)))
```

```
Call:
lm(formula = AUC ~ N.Changes, data = trend_averages, contrasts = list(f = contrasts))

Residuals:
     Min       1Q   Median       3Q      Max 
-0.36910 -0.12677 -0.01689  0.12652  0.36703 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  0.55254    0.01140  48.480  < 2e-16 ***
N.Changes.L  0.12765    0.02279   5.600 5.94e-08 ***
N.Changes.Q  0.01714    0.02279   0.752    0.453    
N.Changes.C  0.02401    0.02279   1.053    0.293    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1766 on 236 degrees of freedom
Multiple R-squared:  0.1228,	Adjusted R-squared:  0.1116 
F-statistic: 11.01 on 3 and 236 DF,  p-value: 8.566e-07
```

Arrays

```
trend_arrays <- sum_data_exp1  %>%
  filter(Representation == "arrays")

contrasts<-contr.poly(4,c(1,2,4,8))
summary(lm(AUC ~ N.Changes, data=trend_arrays, contrasts = list(f = contrasts)))
```

```
Call:
lm(formula = AUC ~ N.Changes, data = trend_arrays, contrasts = list(f = contrasts))

Residuals:
     Min       1Q   Median       3Q      Max 
-0.37902 -0.11908 -0.01044  0.10529  0.37121 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  0.52899    0.01070  49.421  < 2e-16 ***
N.Changes.L  0.10194    0.02141   4.762 3.35e-06 ***
N.Changes.Q  0.03516    0.02141   1.642    0.102    
N.Changes.C  0.01123    0.02141   0.525    0.600    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1658 on 236 degrees of freedom
Multiple R-squared:  0.09802,	Adjusted R-squared:  0.08656 
F-statistic: 8.549 on 3 and 236 DF,  p-value: 2.063e-05
```

Appendix G
R output for Experiment 1: Confidence Analyses

ANOVA and Mauchly's Test for Sphericity

```
options(contrasts=c("contr.sum","contr.poly"))
ezANOVA(data=sum_data_exp1_p, dv=mean_conf, wid= Participant, within= .(Representation, N.Changes))
```

```
$ANOVA
```

	Effect <chr>	DFn <dbl>	DFd <dbl>	F <dbl>	p <dbl>	p<.05 <chr>	ges <dbl>
2	Representation	1	29	17.52347	2.408389e-04	*	0.028725097
3	N.Changes	3	87	43.78755	2.415943e-17	*	0.026629214
4	Representation:N.Changes	3	87	5.57755	1.511730e-03	*	0.002043356

3 rows

```
$`Mauchly's Test for Sphericity`
```

	Effect <chr>	W <dbl>	p <dbl>	p<.05 <chr>
3	N.Changes	0.4882789	0.001331785	*
4	Representation:N.Changes	0.8427013	0.448130640	

2 rows

```
$`Sphericity Corrections`
```

	Effect <chr>	GGe <dbl>	p[GG] <dbl>	p[GG]<.05 <chr>	HFe <dbl>	p[HF] <dbl>	p[HF]<.05 <chr>
3	N.Changes	0.6962842	9.165586e-13	*	0.7514305	1.346408e-13	*
4	Representation:N.Changes	0.9037398	2.254531e-03	*	1.0059625	1.511730e-03	*

2 rows

Polynomial contrasts

```
contrasts<-contr.poly(4,c(1,2,4,8))
summary(lm(mean_conf ~ N.Changes, data=sum_data_exp1_p, contrasts = list(f = contrasts)))
```

```
Call:
lm(formula = mean_conf ~ N.Changes, data = sum_data_exp1_p, contrasts = list(f = contrasts))

Residuals:
     Min       1Q   Median       3Q      Max
-2.45104 -1.24596 -0.09948  0.92786  3.10781

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.095182   0.088955  34.795   <2e-16 ***
N.Changes.L  0.417982   0.177910   2.349   0.0196 *
N.Changes.Q  0.152865   0.177910   0.859   0.3911
N.Changes.C -0.004309   0.177910  -0.024   0.9807
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.378 on 236 degrees of freedom
Multiple R-squared:  0.02583,	Adjusted R-squared:  0.01345
F-statistic: 2.086 on 3 and 236 DF,  p-value: 0.1027
```

Appendix H
R output for Experiment 2: Descriptives

Area Under the Curve:

```
sum_data_exp2_p %>%
  group_by(Representation, N.Images) %>%
  summarise(mean = mean(AUC),
            variance = var(AUC),
            SD = sd(AUC))
```

Representation <fctr>	N.Images <ord>	mean <dbl>	variance <dbl>	SD <dbl>
averages	2 Images	0.9372721	0.007402696	0.08603892
averages	4 Images	0.7891667	0.017492641	0.13225975
averages	8 Images	0.6060742	0.016872239	0.12989318
averages	16 Images	0.4355599	0.015303800	0.12370853
arrays	2 Images	0.8495312	0.019982030	0.14135781
arrays	4 Images	0.6654036	0.024568018	0.15674188
arrays	8 Images	0.4807422	0.022017106	0.14838162
arrays	16 Images	0.4453516	0.024496803	0.15651455

8 rows

Proportion Correct:

```
sum_data_exp2_p %>%
  group_by(Representation, N.Images) %>%
  summarise(
    mean = mean(mean_PC),
    variance = var(mean_PC),
    SD = sd(mean_PC)
  )
```

Representation <fctr>	N.Images <ord>	mean <dbl>	variance <dbl>	SD <dbl>
averages	2 Images	0.9000000	0.007637392	0.08739217
averages	4 Images	0.7937500	0.008930496	0.09450130
averages	8 Images	0.6781250	0.006947064	0.08334905
averages	16 Images	0.5500000	0.005697737	0.07548336
arrays	2 Images	0.8718750	0.008105469	0.09003038
arrays	4 Images	0.7312500	0.009199892	0.09591607
arrays	8 Images	0.6010417	0.006376841	0.07985512
arrays	16 Images	0.5583333	0.010084411	0.10042117

8 rows

Confidence:

```
sum_data_exp2_p %>%
  group_by(Representation, N.Images) %>%
  summarise(
    mean = mean(mean_conf),
    variance = var(mean_conf),
    SD = sd(mean_conf)
  )
```

Representation <fctr>	N.Images <ord>	mean <dbl>	variance <dbl>	SD <dbl>
averages	2 Images	4.270833	1.327205	1.152044
averages	4 Images	3.689583	1.724672	1.313268
averages	8 Images	3.208333	1.888290	1.374151
averages	16 Images	3.006250	1.922912	1.386691
arrays	2 Images	4.434375	1.745174	1.321051
arrays	4 Images	3.413542	2.117840	1.455280
arrays	8 Images	2.806250	1.891056	1.375157
arrays	16 Images	2.521875	1.640433	1.280794

8 rows

Appendix I
R output for Experiment 2: Discriminability Analyses

t-test

```
t.test(
  sum_data_exp2$AUC[sum_data_exp2$Mode == "sim"],
  sum_data_exp2$AUC[sum_data_exp2$Mode == "participant"],
  paired = TRUE,
  alternative = "two.sided"
)
```

```

    Paired t-test

data:  sum_data_exp2$AUC[sum_data_exp2$Mode == "sim"] and sum_data_exp2$AUC[sum_data_exp2$Mode == "participant"]
t = -11.278, df = 239, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.2077574 -0.1459714
sample estimates:
mean of the differences
             -0.1768644
```

Hide

```
t.test(
  sum_data_exp2$mean_PC[sum_data_exp2$Mode == "sim"],
  sum_data_exp2$mean_PC[sum_data_exp2$Mode == "participant"],
  paired = TRUE,
  alternative = "two.sided"
)
```

```

    Paired t-test

data:  sum_data_exp2$mean_PC[sum_data_exp2$Mode == "sim"] and sum_data_exp2$mean_PC[sum_data_exp2$Mode == "participant"]
t = -16.95, df = 239, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.2194659 -0.1737633
sample estimates:
mean of the differences
             -0.1966146
```

ANOVA and Mauchly's Test for Sphericity

```
options(contrasts=c("contr.sum","contr.poly"))
ezANOVA(data=sum_data_exp2_p, dv=AUC, wid= Participant, within= .(Representation, N.Images))
```

```
$ANOVA
```

	Effect <chr>	DFn <dbl>	DFd <dbl>	F <dbl>	p <dbl>	p<.05 <chr>	ges <dbl>
2	Representation	1	29	17.051560	2.812690e-04	*	0.08539299
3	N.Images	3	87	273.164590	3.782164e-44	*	0.62724821
4	Representation:N.Images	3	87	6.165894	7.529199e-04	*	0.04047322

3 rows

```
$`Mauchly's Test for Sphericity`
```

	Effect <chr>	W <dbl>	p <dbl>	p<.05 <chr>
3	N.Images	0.7747150	0.2153003	
4	Representation:N.Images	0.8645411	0.5446119	

2 rows

```
$`Sphericity Corrections`
```

	Effect <chr>	GGe <dbl>	p[GG] <dbl>	p[GG]<.05 <chr>	HFe <dbl>	p[HF] <dbl>	p[HF]<.05 <chr>
3	N.Images	0.8432236	1.301027e-37	*	0.9304773	2.990786e-41	*
4	Representation:N.Images	0.9242339	1.080727e-03	*	1.0317628	7.529199e-04	*

2 rows

Polynomial contrasts

Averages

```
trend_averages <- sum_data_exp2  %>%
  filter(Representation == "averages")

contrasts<-contr.poly(4,c(2,4,8,16))
summary(lm(AUC ~ N.Images, data=trend_averages, contrasts = list(f = contrasts)))
```

```
Call:
lm(formula = AUC ~ N.Images, data = trend_averages, contrasts = list(f = contrasts))

Residuals:
     Min       1Q   Median       3Q      Max 
-0.41681 -0.13921 -0.00658  0.16053  0.36175 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  0.58352    0.01199  48.651  < 2e-16 ***
N.Images.L  -0.17941    0.02399  -7.479 1.46e-12 ***
N.Images.Q  -0.01165    0.02399  -0.486    0.628    
N.Images.C   0.01550    0.02399   0.646    0.519    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1858 on 236 degrees of freedom
Multiple R-squared:  0.1934,	Adjusted R-squared:  0.1832 
F-statistic: 18.86 on 3 and 236 DF,  p-value: 5.308e-11
```

Arrays

```
trend_arrays <- sum_data_exp2  %>%
  filter(Representation == "arrays")

contrasts<-contr.poly(4,c(2,4,8,16))
summary(lm(AUC ~ N.Images, data=trend_arrays, contrasts = list(f = contrasts)))
```

```
Call:
lm(formula = AUC ~ N.Images, data = trend_arrays, contrasts = list(f = contrasts))

Residuals:
     Min       1Q   Median       3Q      Max 
-0.40588 -0.10668 -0.01805  0.11253  0.35059 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  0.54189    0.01064  50.935  < 2e-16 ***
N.Images.L  -0.14661    0.02128  -6.890 5.02e-11 ***
N.Images.Q   0.04274    0.02128   2.009   0.0457 *  
N.Images.C   0.02826    0.02128   1.328   0.1854    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1648 on 236 degrees of freedom
Multiple R-squared:  0.1842,	Adjusted R-squared:  0.1738 
F-statistic: 17.76 on 3 and 236 DF,  p-value: 1.991e-10
```

Appendix J
R output for Experiment 2: Confidence Analyses

ANOVA

```
options(contrasts=c("contr.sum","contr.poly"))
ezANOVA(data=sum_data_exp2_p, dv=mean_conf, wid= Participant, within= .(Representation, N.Images))
```

```
$ANOVA
```

	Effect <chr>	DFn <dbl>	DFd <dbl>	F <dbl>	p <dbl>	p<.05 <chr>	ges <dbl>
2	Representation	1	29	4.826753	3.616278e-02	*	0.008969519
3	N.Images	3	87	105.050638	7.978797e-29	*	0.177547640
4	Representation:N.Images	3	87	9.515248	1.679894e-05	*	0.008979358

3 rows

Mauchly's Test for Sphericity

```
$`Mauchly's Test for Sphericity`
```

	Effect <chr>	W <dbl>	p <dbl>	p<.05 <chr>
3	N.Images	0.2178271	5.437667e-08	*
4	Representation:N.Images	0.4778507	1.028516e-03	*

2 rows

```
$`Sphericity Corrections`
```

	Effect <chr>	GGe <dbl>	p[GG] <dbl>	p[GG]<.05 <chr>	HFe <dbl>	p[HF] <dbl>	p[HF]<.05 <chr>
3	N.Images	0.5587653	3.798243e-17	*	0.5890963	5.955917e-18	*
4	Representation:N.Images	0.7078770	1.891281e-04	*	0.7653431	1.171868e-04	*

2 rows

Polynomial contrasts

```
contrasts<-contr.poly(4,c(2,4,8,16))
summary(lm(mean_conf ~ N.Images, data=sum_data_exp2_p, contrasts = list(f = contrasts)))
```

```

Call:
lm(formula = mean_conf ~ N.Images, data = sum_data_exp2_p, contrasts = list(f = contrasts))

Residuals:
    Min      1Q  Median      3Q     Max 
-3.3214 -1.0219 -0.0797  0.8888  3.2359 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)  3.418880   0.086210  39.658  < 2e-16 ***
N.Images.L  -1.187329   0.172420  -6.886 5.14e-11 ***
N.Images.Q   0.278906   0.172420   1.618    0.107    
N.Images.C   0.009899   0.172420   0.057    0.954    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.336 on 236 degrees of freedom
Multiple R-squared:  0.1749,	Adjusted R-squared:  0.1645 
F-statistic: 16.68 on 3 and 236 DF,  p-value: 7.324e-10
```