The visual system encounters an enormous amount of complex information that is processed to produce a smooth phenomenal experience of the world. The visual processes that achieve this remarkable feat require a memory store that encodes, retains and manipulates the visual information. For example, an active memory store integrates the information between saccades [@irwin_integration_1996], orients where attention should be deployed [@awh_overlapping_2001], and retains information about objects during visual tracking and search [@carlisle_attentional_2011]. The system responsible for actively storing the visual information for perception has been termed "visual working memory" (VWM). Despite its necessity in everyday perception, the VWM system is surprisingly limited in the amount of information it can encode, approximately three to four objects [@luck_capacity_1997]. This thesis explores the processes that contribute to this capacity limit with research that examines how memory performance can be boosted to overcome this limit. This chapter provides an overview of past visual working memory research.

The conception of working memory

Classical research separated memory into two distinct but interacting systems, short-term memory (STM) and long-term memory (LTM). The STM store has a highly limited capacity that holds current information in awareness, whereas LTM is thought to be unlimited in capacity, but the information stored is effortfully retrieved [@atkinson_human_1968]. @atkinson_human_1968 were one of the first to consider the STM system as "working"; "a system in which decisions are made, problems are solved and information flow is directed". That is, working memory functioned as a mental work space for higher-level cognition [@nee_working_2018]. However, this early conception of STM as "working" relied on the incorrect assumption that encoding of information into LTM, and therefore learning, required repeated maintenance in STM, which has since been shown to be untrue [@baddeley_working_1974]. This was updated by Baddeley and Hitch's [-@baddeley_working_1974] highly influential multi-component working memory model. Their first iteration contained three subsytems: the central executive, the phonological loop and the visuospatial sketchpad (Figure 1). The phonological loop and the visuospatial sketchpad, collectively known as the "slave systems", maintain verbal and visual information respectively. The visuospatial sketchpad is analogous to what researchers now refer to as the VWM system.
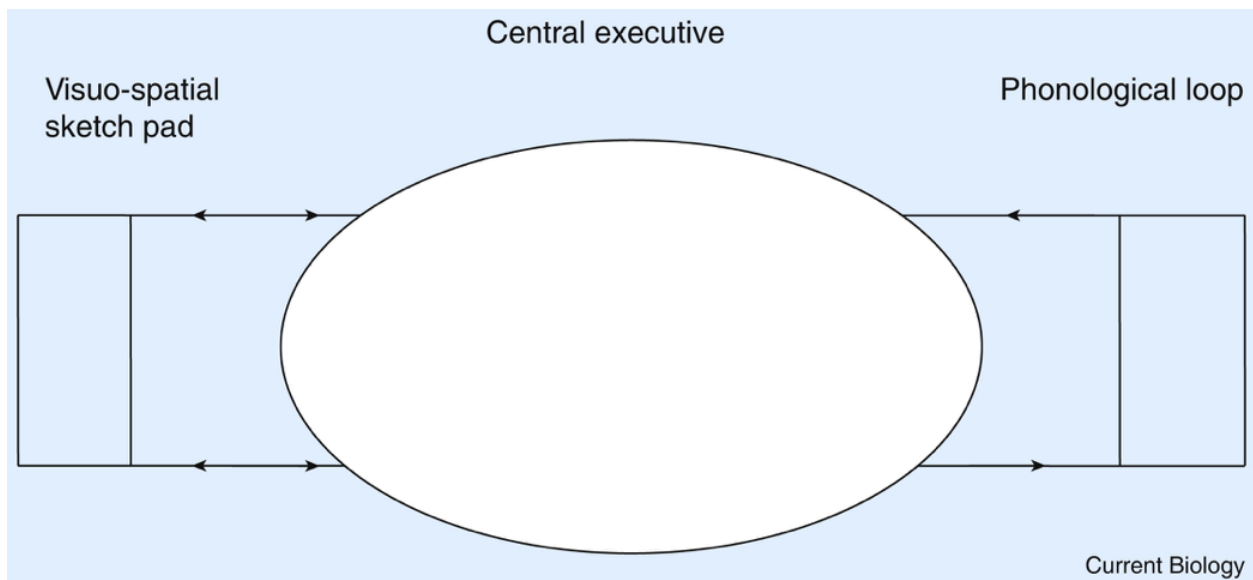

Figure 1: An early model of working memory proposed by Baddeley. Figure taken from @baddeley_working_1974.

Baddeley and Hitch's [-@baddeley_working_1974] model provided key foundations that defined working memory [@nee_working_2018]. Firstly, the processes involved in the temporary maintenance of information

are distinguishable from those involved in permanent retention of information into long-term memory. Secondly, the processes that modulate and manipulate the retained information are dissociable from processes that only retain the information, such as those involved in iconic memory. Thirdly, the memory processes are modal such that visual materials are represented differently to verbal materials.

Performance in WM tasks has since been shown to correlate robustly with various measures of cognitive ability and intelligence [@daneman_individual_1980, @unsworth_working_2014]. [@cowan_capacity_2005, @fukuda_quantity_2010]

While the term "*visual working memory*" is often used synonymously with "*visual short-term memory*", @luck_visual_2013 provides three defining aspects of VWM: the information represented is visual in nature, VWM information is actively maintained and that the information is accessed for cognitive use.

Measuring visual working memory capacity

In their seminal study, @luck_capacity_1997 devised the change-detection paradigm for the measurement of VWM capacity. In this paradigm (see Figure 2), an initial array (*sample array*) of objects is presented to the observer for a brief duration, usually no longer than a second, before disappearing. After a short delay, a second array (*test array*) may appear identically to the sample array (*no-change* trials) or with one object replaced by another object (*change* trials). The observer has to respond with whether they think a change occurred or not on that trial.
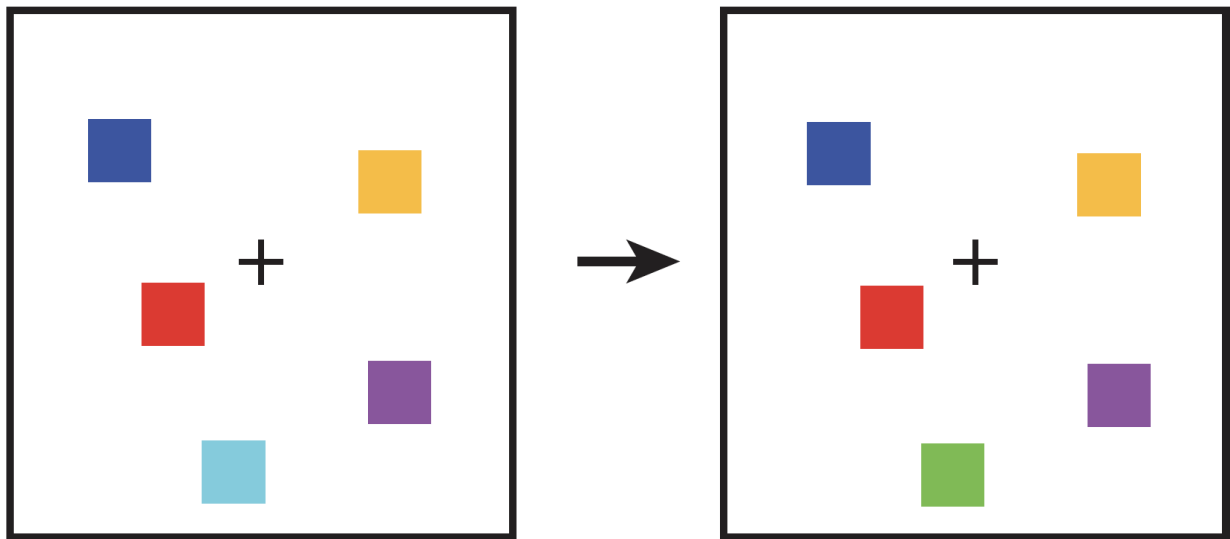

Figure 2: An example of what is displayed on a change-detection trial.

The proportion of trials that a participant correctly detects a change or no change occurred can be used to estimate the number of items held in visual working memory. Assuming the observer has stored a certain number of objects (K) from the sample array, a correct response on a change trial (a 'hit') will occur whenever the changed item is one of those K objects. If an array contains N objects, on average this will occur on K out of N change trials. Additional hits will occur on a porportion (G) of the remaining (N-K) out of N change trials (when the changed object is not among those encoded) if the observer correctly guesses that a change has occurred. For an unbiased observer, this will occur on half of the remaining trials ($G = 0.5$), but G can be estimated using the observer's false alarm rate, the overall number of trials which a change is reported when no change occurred. This produces the model proposed by @pashler_familiarity_1988:

$$H = \frac{K}{N} + \frac{(N-K)}{N} * G \tag{1}$$

where H is the probability of a hit on a change trial. Rearranged to make K the subject:

$$K = \frac{N * (H - G)}{1 - G} \tag{2}$$

However, this equation assumes VWM has no bearing on a no-change trial [@cowan_capacity_2005]. On no-change trials, the guesswork is limited to items not stored in VWM (*N-K*). Cowan estimates that the subject will that a change has not occurred with a probability of *1 - G*, where *G* is the probability of guessing a change had occurred. This was updated by @cowan_metatheory_2001 to include the correct rejection rate (*CR*):

$$CR = \frac{K}{N} + \frac{(N - K)}{N} * (1 - G) \tag{3}$$

Adding this to Equation 1:

$$H + CR = \frac{2K}{N} + \frac{(N - K)}{N} = \frac{(K + N)}{N} \tag{4}$$

Rearranging to make K the subject:

$$K = N * (H + CR - 1) \tag{5}$$

The capacity of visual working memory

Despite its necessity, the capacity of visual working memory is surprisingly limited to approximately 3-4 items' worth of information. @luck_capacity_1997 presented sample arrays containing from 1 to 12 coloured squares for 100ms, before showing the test array approximately a second later. They found performance was almost perfect for arrays of 1 to 3 colour blocks and declined from 4 to 12 colour blocks. This pattern remained when observers were given two digits to rehearse aloud to suppress any influence of verbal working memory (see Figure 3a), when the sample duration was displayed for a longer duration, and when observers were required to only make a decision about a single cued item in the array (see Figure 3b). Estimating VWM capacity from the change-detection accuracy (see Equation 2) indicated observers stored approximately four items in VWM.

Despite agreement of this capacity limit for simple visual objects, there has been contention over the architecture of VWM producing this limit. In addition to simple colours, @luck_capacity_1997 increased the number of relevant features in the visual stimuli presented in the same change-detection task and found an identical pattern of memory performance when presenting colours. For example, with conjunctions of colour and orientation, VWM performance was no different when instructed to detect changes in only colour, only orientation or in either feature (see Figure 3c). This pattern was also replicated with stimuli that were conjunctions of four features: colour, orientation, size and the presence of a gap (see Figure 3d) and conjunctions of the same feature type, such as two colours (see Figure 3e). Since increasing the number of relevant features in the visual stimuli did not influence memory performance, @luck_capacity_1997 proposed that the architecture of VWM is 3 to 4 'slots' where each slot stores a representation of the visual object with its features integrated, rather than the individual features of the object.

The 'slots' model was directly opposed by the findings of @alvarez_capacity_2004. In their study, participants completed the same change-detection task as in @luck_capacity_1997 but with different stimulus sets. The stimuli sets included colour squares as @luck_capacity_1997 had done, but also Snodgrass line drawings, shaded cubes, random polygons, Chinese characters and English letters (see Figure 4). VWM capacities were significantly different for the stimulus sets contradicting what would be predicted by the 'slots' model.
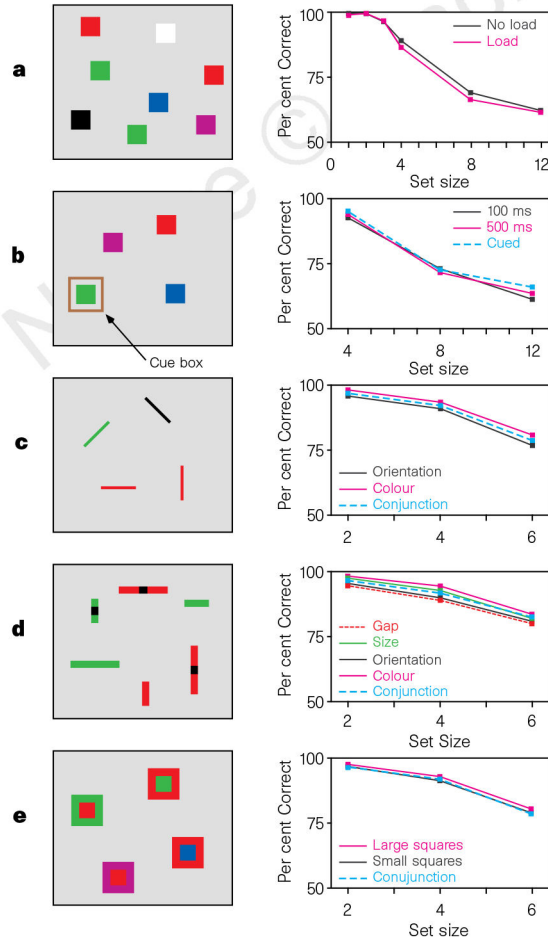
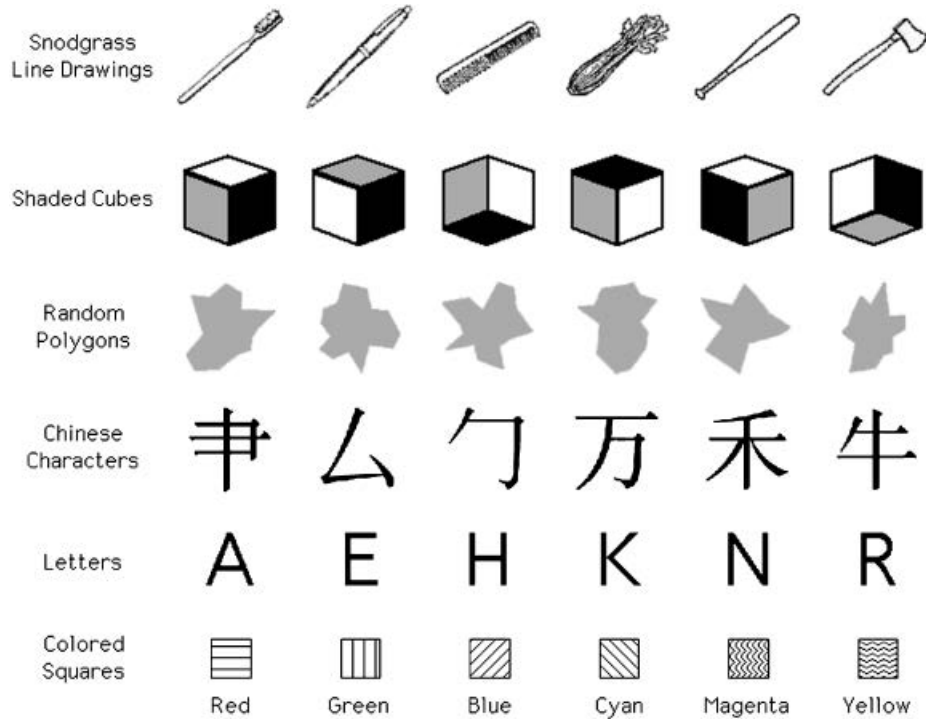Figure 3: Stimulus arrays and memory performance from multiple experiments in @luck_capacity_1997.

Figure 4: The stimuli sets used in Alvarez and Cavanagh (2004)

Critically, @alvarez_capacity_2004 indexed the *complexity* of each stimulus set by conducting a visual search task with the same stimulus sets. In the visual search task, observers were presented a target object before asking to locate whether that target was present in an array of objects from the same stimulus set. The arrays contained either 4, 8, or 12 objects and included the target object on half the trials. The *visual search rate*, their measure of stimulus complexity, was the estimated amount of additional reaction time taken to respond that the target was present with each additional item in the array. Estimating capacity as the number of objects for each stimulus set that would correspond to 75% accuracy on the change-detection task, visual search rate was very strongly correlated ($r = .992$) to the inverse of capacity (see **??**).

@luck_capacity_1997 and @alvarez_capacity_2004 provide contrasting findings. While @luck_capacity_1997 show VWM capacity is consistently approximately 4 objects when varying the number of features being combined, whereas @alvarez_capacity_@2004 find VWM capacity is different for stimuli of varying complexity. @alvarez_capacity_2004 suggested VWM capacity is limited by total amount of visual information rather than the number of objects as @luck_capacity_1997 suggested in their 'slots' model. They posited the 'resource' model, which suggests that more complex visual items (those with more features) require more resources to be encoded and stored. Thus, as the visual stimuli get more complex, less items are maintained in VWM.

@awh_visual_2007 disputes whether the varying VWM capacities found by @alvarez_capacity_2004 was due to stimulus complexity. They suggest that the variation in VWM performance is due to an increase in comparison errors made when the object stored in memory is visually similar to the object that changed in the test array. To examine this, @awh_visual_2007 gave participants a change-detection task with memory arrays containing 4 or 8 items selected from a stimulus set of 6 shaded cubes and 6 Chinese characters. This meant that either a *within-category* change would occur, where a shaded cube changed to another shaded cube or a Chinese character changed to another Chinese character, or a *cross-category* change would occur, where a shaded cube would change to a Chinese chracter or a Chinese character would change to a shaded cube. A *within-category* change is more likely to produce more errors as the to-be compared items come from the same stimulus set, whereas the to-be compared items in a *cross-category* change come from the other stimulus set and are therefore, relatively dissimilar. If stimulus complexity influences the number of items
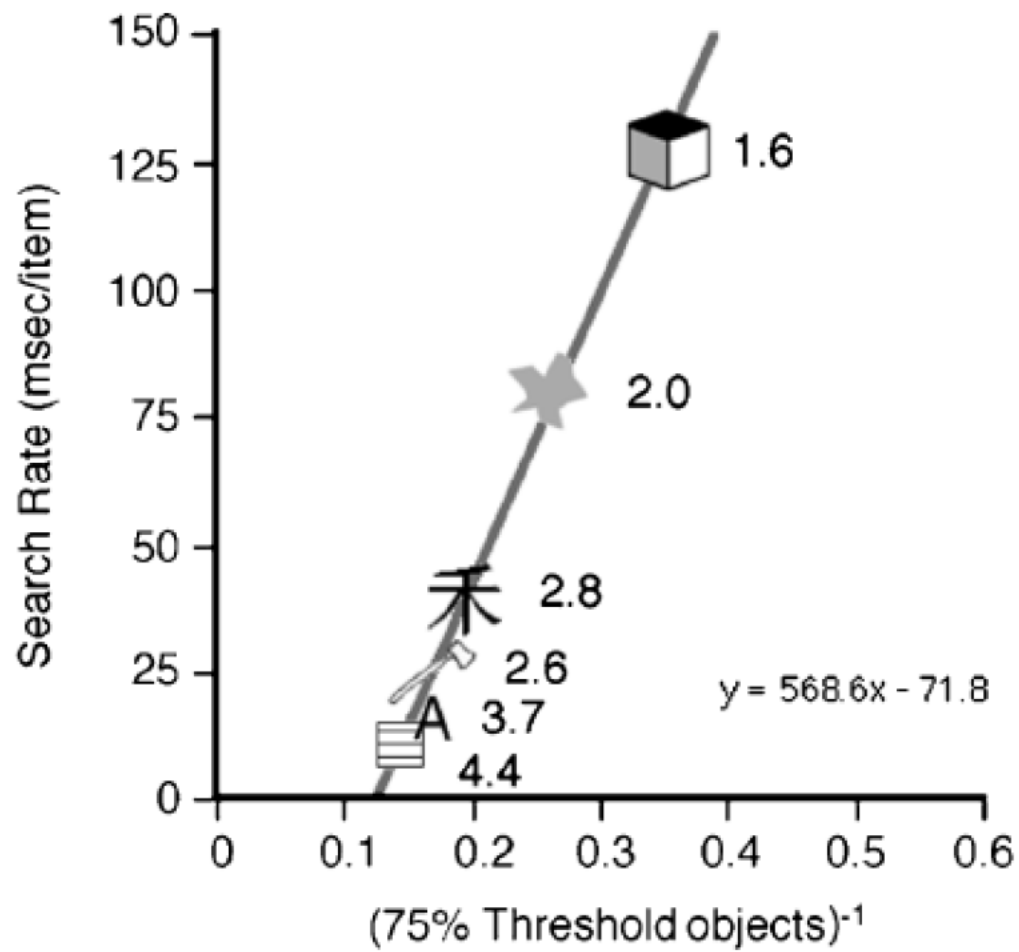
Figure 5: The visual search rate is highly correlated to the number of objects that corresopnds to 75% accuracy on the change-detection task. The values beside each stimulus item is the calculated capacity for each stimulus set. Taken from Alvarez and Cavanagh (2004)

stored in VWM, then there should be no benefit of a *cross-category* change compared to a *within-category* change. But if stimulus complexity influences the comparison decision, there should be an improvement in performance for *cross-category* changes relative to *within-category* changes. They found that memory performance for *within-category* changes was significantly worse, although significantly worse for a Chinese character compared to a shaded cube. Additionally, change-detection performance for *cross-category* was equivalent for change-detection performance with colours, which has relatively low *sample-test similarity*.
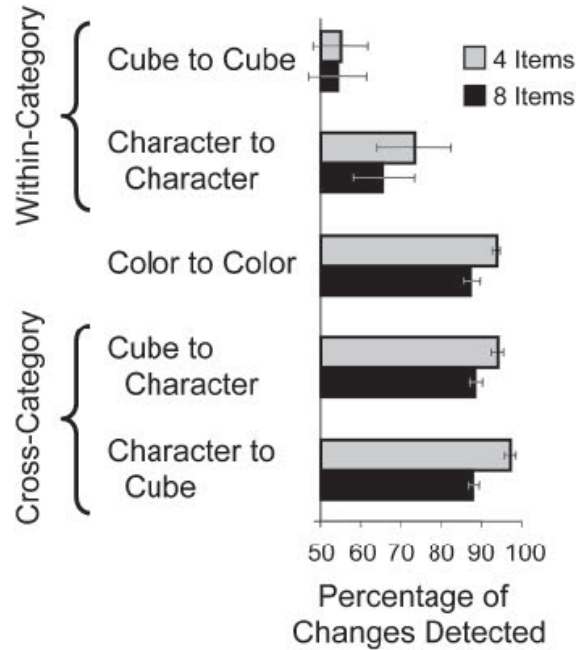

Figure 6: Results of Experiment 2 from Awh et al. (2007), showing that change-detection accuracy was significantly better for cross-category changes compared to within-category changes.

From this, @awh_visual_2007 concluded that the number of items represented in visual working memory is fixed, regardless of the complexity of those items. However, their findings did not contradict the basic conclusion of @alvarez_capacity_2004 that stimulus complexity does influence change-detection performance. A key insight from the @awh_visual_2007 paper is that rather than the stored number of items, it may be the resolution with which objects are stored in visual working memory may be the key limiting factor in change-detection performance. That is, more complex objects are stored with a limited resolution such that it is more difficult to detect *within-category* changes, leading to poorer overall change-detection performance at the same set size.

The precision of representations in visual working memory

Current models of visual working memory

However, the effect of training participants to be familiar with stimuli on visual working memory performance is unclear. To train recognition to polygons, Chen, Eng and Jiang (2006) presented four polygons out of a training set of eight, before presenting two polygons, one the same and one from the unpresented set. Despite being able to recognise the trained polygon, this familiarity did not improve visual working memory performance for the trained polygons over novel polygons. However, Blalock (2015) found a positive effect of familiarity training on visual working memory performance. Blalock (2015) presented a target polygon before asking the participant to select the target out of an array of four polygons. This recognition training produced better change-detection performance for trained polygons over the novel polygons. Another notable discrepancy between these studies is the sample size. While Chen et al. (2006) used twelve participants in

each of their experiments, Blalock (2015) used over seventy and 102 in each of theirs. This difference in the statistical power of experiments may explain the contrasting results of familiarity training.