2

A Review of Literature

Visual feature binding is a process, which accumulates all the distinct features of a visual stimulus and creates a unified representation of it. Various features of an object such as color and shape are initially processed to reach separate specialized areas of the brain. Despite this scattered information, how one encodes a unified representation of a visual stimulus and maintains the encoded representation, is the unsolved binding conundrum. Researchers have attempted to solve the binding problem for several decades, using a variety of approaches – psychological, biological, computational, and philosophical. The difference is more in the primary perspective taken, rather than the ideas about feature binding, as researchers often borrow concepts, ideas, and evidences generated by people from a different approach. Keeping this in mind, this chapter begins with a review of ideas regarding the biological processes involved in feature binding. Thereafter, the major variables in this study are reviewed as the important psychological factors in feature binding along with their neural substrates. Then, a focused review of research studies assessing the brain areas involved in feature binding, particularly as a result of the major variables under study is presented. Towards the end of the chapter, the general objectives of the present research work are delineated.

2.1 FEATURE BINDING

Many researchers believe that the answer to the 'problem of binding' lies in identifying dedicated neurons or networks that are involved in the binding event or process. Evidence for conjunctively coding cells that responded to whole objects in the visual cortex of cats and monkeys [Hubel and Wiesel, 1959; 1962; 1968], was so exciting and important that Hubel and Wiesel were awarded the Nobel Prize in 1981 for their discoveries regarding "information processing in the visual system". From their work, it was clear that specialized neurons existed for different objects in the environment. Such cells were christened the 'grandmother cell', as presumably people possessed a cell for coding their grandmother, a different one for their mother, a still different one for their car, and so on. The premise underlying the grandmother cell was that for every complex stimulus there is one particular cell, which is involved in the representation of that stimulus. Gross [2002] who has reviewed the history of research on grandmother cells, states that it was Lettvin, who first used the term "grandmother cell" in 1967 in a parable. Barlow [1995] also mentions Lettvin as the originator of the notion of the grandmother cells to refer to neurons encoding highly specific information. Gross [2002] also discusses that unknown to Lettvin, Konorski [1967] had also experimented with a similar concept, terming such nerve cells as 'gnostic' neurons. Konorski [1967] gave a conceptual map of the visual cortex, which represents gnostic neurons for specific stimuli in different areas. The areas, which contain gnostic neurons, are termed 'gnostic fields'. Such conceptualizations supported the 'grandmother cell' idea.

An initial problem was how such presumably permanent cells contend with the fact that binding is transient and flexible, i.e., what is bound can also be unbound. In response, many researchers proposed that specialized cells initially respond to specific conjunctions or bindings and further processing happens in a "workspace" that allows somewhat more flexibility in binding. Fodor and Pylyshyn [1988] differentiated "modular faculties" and a "central horizontal system" which traversed modules to redistribute information. Baars [1988] proposed several unconscious, specialized, parallel processors, on one hand, and a single, limited capacity, serial workspace that permits exchange of information, on the other hand. Deheane and Changeux [2004] postulated the "neuronal workspace hypothesis", which distinguished a network of "specialized processors" allowing automatic, quick feed-forward processing, and "cortical workspace neurons" which transcend the modules of the cortex, as they were long range excitatory neurons that could send signals to and receive signals from many disparate areas in the brain.

The idea of specialized neurons being responsible for feature bindings also had a problem with sheer numbers. Riesenhuber and Poggio [1999] criticized the idea of grandmother cell on this account. They argued that the number of complex stimuli in the world is infinite and it is impossible that the visual cortex could have a cell reserved specially for each stimulus in the world. The vexing problem really was that there were numerous stimuli, and these were also bound only in a transient way (compounding the numerical issue); whereas the number of conjunctively coding neurons was finite and relatively limited. Gradually, the idea of 'grandmother cells' was replaced with several neurons together being dedicated to specific kinds of objects. Among many others, Gross [2002] discussed that there are hand and face selective neuronal cells residing in the inferior temporal cortex. Quiroga et al. [2005] suggest an invariant, sparse, and explicit coding by neurons in the medial temporal lobe [MTL], which could be important in the transformation of complex visual percepts into long-term and more abstract memories. Quiroga et al. [2008] reviewed the evidence from a large number of neuropsychological and imaging studies that consistently found that MTL neurons respond selectively to complex visual stimuli and suggested that these neurons could be considered as 'grandmother cells', although they exhibited a sparse and abstract representation of visual information. Based on these characteristics, they further contended that these neurons are not really so specific, and are amenable to training/ learning/ plasticity. E.g., Thome et al. [2012] found that even passive viewing for different periods led to differences in the tuning properties of MTL neurons in monkeys. A recent set of articles [Barlow, 2009; Quiroga, 2012; Reddy and Thorpe, 2014; Roy, 2013, 2017, 2020] suggests that the idea of abstract, invariant representations coding specific kinds of information is alive and well.

Synchrony was the second physiological explanation for feature binding. Among the pioneers, Von der Malsburg [1981] argued that a complex environment comprising different objects or events requires parallel processing of information, and proposed neural synchrony as the mechanism for binding. Gray et al. [1992] found transient and precise synchronization of neurons in the cat striate cortex. Singer and Gray [1995] proposed that this was the mechanism responsible for feature binding. Synchronization is found in species as varied as locusts [MacLeod et al., 1998], monkeys [Gray, 1999], and human beings [Singer, 1999]. Neural synchrony implied that binding probably occurs due to activation all over the brain, with synchronous firing of neurons in the cortex being associated with binding of separate features of an object, which are represented in different areas of the brain. Currently, it is the most popular and parsimonious explanation of feature binding [see recent review by Ding et al., 2017].

The idea of synchrony was not without its critics. Vociferous detractors alleged that, "the theory is inadequate in conception and impoverished in support" [Shadlen and Movshon, 1999, p. 77]. They contended that it merely describes the correlates of binding without saying anything about how bindings are caused. Moreover, synchrony can be found only at the very

early stages of cortical processing, whereas both neurological and behavioral results of perceptual binding suggest that it must be a high-level computation. Considering the architecture of the cerebral cortex, they suggest that it lacks the mechanisms needed to decode synchronous spikes and to treat those underlying separate objects as a special code. Finally, they contend that there is experimental evidence for correlated cortical activity, but there is very little that directly or compellingly links this activity to binding. Roelfsema *et al.* [2004] also criticized the synchrony explanation of binding when they contended that neural synchrony is not always correlated with contour detection or perception, which seems necessary for segregation of the object from the background.

Primarily, the idea of synchrony faces two problems. The first problem is how two [or more] objects are differentiated. In response, it was proposed that the organism rapidly shifts from one to other set of neurons, which fire out of phase when they are encoding multiple objects. But, it is difficult to imagine the implementation of this precise timing mechanism in the brain, particularly with multiple objects in the external world, and the large levels of background noise in the brain. The second problem is that if the same neurons encode all binding operations, how is it possible to have permanent or durable representations. Synchrony is evidently a good explanation of binding, but only for a single object, and that too, only when it is perceived in the immediate environment. It does not explain memory for objects, which are no longer present.

As mentioned in the preceding paragraphs, research evidence is abundant, and in fact exists across species, for specialized neurons as well as synchrony. Some researchers tried to resolve the debate between synchronously firing neurons and specialized neurons by proposing different types of bindings. Crick and Koch [1990] distinguished between three kinds of bindings. The first type was bindings 'hardwired' genetically to determine the response to natural stimuli. The second type of bindings were learnt due to experience such as those required for recognizing familiar faces, or the alphabet; and the third type were transient bindings of novel stimuli, which require focused attention. The transient bindings are the ones primarily based on neural synchrony, and develop into the second kind with the frequent repetition of the stimulus. Baddeley [2007] suggested a distinction between passive binding and active binding. Passive binding refers to binding natural world elements for which humans are 'prepared' by evolution, and is thus automatic. On the other hand, active binding requires attention as it is the binding of arbitrary, learnt elements. He also opined that long-term episodic memory may be another source of binding. VanRullen [2009] also distinguished between hardwired automatic binding of natural/frequently encountered objects, and on-demand binding of meaningless/arbitrary feature conjunctions, asserting that while the latter always require attention, the former require attention only if there is competition by multiple objects. Hommel and Colzato [2009] explicitly acknowledged two binding mechanisms - neural synchronization of all features present at a time, or a permanent detector for real/familiar objects. They also suggested that both presumably act together as the organism navigates a dynamic yet stable environment.

In spite of such resolutions to the debate between specialized neurons and synchrony, there is also a realization that both these explanations conceptualize binding only as a fast feed forward mechanism. This does not sit well with processing of objects in reality. An object is not merely a bundle of features. Nor is separation of an object from the background enough for extracting meaning from it. At the very least, the object needs to be distinguished from other objects. This implies some selection and/or manipulation of the rudimentary information brought into the system by feed forward processes.

At the physiological level, this is possibly instantiated through the reentrant processes in the brain. Reentrant processes are implemented in the downward and lateral connections that work to give feedback to lower levels in the brain. For example, the lower level Area V1 in the

visual system sends signals for forward processing, but all higher visual areas also have reentrant [downward] connections with Area V1. Such reentrant connections not only feedback to the original neurons, but also 'receive' signals back from them. Thus, there is continuous communication between brain areas. A preprint recently uploaded by researchers at the University of Cambridge [Zamboni *et al.*, 2020] documents their use of ultra-high field [UHF] imaging to examine blood oxygen level dependent [BOLD] signals across different levels of cortical depth. They demonstrate that visual adaptation [to repeatedly presented gratings with the same orientation] was implemented by suppressive local recurrent processing within the visual cortex, as indicated by stronger BOLD signal decrease in superficial, rather than middle and deeper layers. Further, functional connectivity analysis showed two different mechanisms for adaptive processing: feed-forward connectivity within the visual cortex, and feedback connectivity from posterior parietal to visual cortex, reflecting top-down influences due to expectations for repeated stimuli.

Not surprisingly, re-entrant processes are also linked to higher cognitive functions. Edelman [1978] was arguably the first person to hypothesize that reentrant signaling might be important to integrate activity in disparate cortical areas, and underlie higher brain functions. Damasio [1989] held that memory involves a reactivation of the areas involved at the time of the initial registration of bindings. In his words, 'convergence zones' in the brain enable retro-activation of multiple regions.

The basic premise of the reentrant theory is that brain processes are inherently iterative because in the higher areas of the brain, the receptive fields become larger and lose their feature specificity. Thus, stable representations can be established only by cycling the process in the hierarchical system. In this way, reentrant processes help to confirm the representation, resolve competition, thereby allowing accurate bindings to take place [Bullier, 2001; Di Lollo *et al.*, 2000; Hamker, 2003; Hochstein and Ahissar, 2002, Lamme and Roelfsema, 2000]. Recently, Bouchacourt and Buschman [2019] have presented a flexible model of working memory showing how structured sensory networks and a flexible, higher-order network can work together through recurrent processes to explain the two key characteristics of working memory – the limited capacity and flexible use of resources when more precision is required. Adam and Serences [2019] have referred to this model as the 'coordinated network model'.

In the visual system, reentrant signals probably originate in the parietal cortex. Saalmann *et al.* [2007] held that top-down feedback from the parietal cortex to early sensory areas aided monkeys to focus attention on relevant features in carrying out the visual matching task in the experiment. Silvanto *et al.* [2009] used trans cranial magnetic stimulation [TMS] pulses over the posterior parietal cortex [PPC] in human participants, to demonstrate how they resulted in excitation in the visual areas with one-sided application. This confirmed top-down modulation of visual areas by the posterior parietal cortex among humans.

Reentrant connections in the brain may also be combined with changes in synchronous activity of neurons in the early stages to explain how an object is segregated from the background, and also from other objects [Seth *et al.*, 2004; Van der Togt *et al.*, 2006]. Reentrant processes can also combine with dedicated neural circuits such as those showing conjunctive coding [Dresp-Langley, 2012] Thus, reentrant processes are now accepted to be crucial for accurate binding.

These substantial evidences have resulted in a reconsideration of the concept of binding. Several researchers no longer view binding as only a one-time event, but have tried to describe the stages or phases of the binding process.

Treisman [1996] proposed three sequential mechanisms to solve the binding problem: selection of particular locations by a spatial attention, inhibition of locations from feature maps containing unwanted features, and top-down activation of the location containing the currently

attended object for further processing. She also speculated that reentry to area V1 or V2 mediated all these three different mechanisms, proposing that reentrant connections from parietal areas mediate spatial attention; from extra striate areas mediate feature-based selection; and from the inferior temporal cortex mediate object-based selection. Treisman [2006] holds that the initial response of the brain is to activate feature detectors in the early striate and extra striate areas that automatically connect to compatible temporal lobe object nodes, and perhaps inhibit the conflicting ones. Parietal cortex then controls a serial reentry scan of the V1 and V2 areas to retrieve the features present in each, and then these are combined to form integrated object representations or bindings.

Humphreys [2001] also proposed a two-stage account of binding, in which forms appear earlier, and then they are filled with surface features, such as colors. The initial evidence for this two-stage process came from a patient GK with bilateral parietal lesions [Humphreys *et al.*, 2000]. The patient could bind form elements into shapes, but could not integrate shapes with color. This led to the idea that the initial stage of binding results in shapes, and thereafter, surface features are associated with the shapes. Interestingly, using fMRI, Papale *et al.* [2019] recently showed that shape itself is processed due to joint contributions of different areas in visual cortex tuned to components of shape such as dots, lines, curvatures, etc. Cinel and Humphreys [2006] held that visual elements are not very strongly bound in the initial 'noisy' stage. These bindings dissipate unless they consolidate into stable and stronger representations after being reinforced by 'top-down attentional feedback' coming from the posterior parietal cortex. Braet and Humphreys [2009] noted more conjunction errors in a patient having bilateral parietal lesions. They also applied TMS to the posterior parietal cortex of normal participants. This increased conjunction errors 150-200 ms after stimulus onset, supporting that binding occurred due to reentrant communication emanating from the posterior parietal cortex.

Roelfsema [2006] postulated two mechanisms responsible for binding in the visual system – 'base-grouping' and 'incremental grouping'. 'Base-groupings' are coded by single dedicated neurons tuned to multiple features, and process selected information through feed-forward connections. But, the innumerable feature combinations in the world cannot be coded this way. Thus, a more flexible form of grouping, called 'incremental grouping' was proposed, which follows base grouping. Incremental grouping enhances the response of all the neurons, which encode separate features bound in perception. It requires more time than base grouping because it relies on horizontal and feedback connections, besides the feed-forward ones. At the behavioral level, the enhancement of the firing rate of certain neurons during incremental grouping is associated with attention directed to features, such that those features are bound together. A computational model predicting figure-ground separation and binding, using base and incremental grouping has also been proposed [Jehee *et al.*, 2007].

Zimmer *et al.* [2006] distinguished transient and durable bindings, implying that a process of consolidation transforms the former into the latter. Murre *et al.* [2006] also distinguish between 'transient' and 'permanent' binding. The former reflects the capacity of WM to select task relevant information for processing, whereas the latter is the capacity of the neural system to store meaningful patterns in long term memory [LTM]. There is a constant interaction between these two. "What is transiently bound in WM governs what is temporarily and eventually permanently bound in long term memory. In turn, what is permanently bound affects transient binding in WM. The interplay of these binding processes determines how the brain develops into a structured system that is cumulatively correlated with its environment, thus implementing a process that is able to lift itself to higher levels of cognitive functioning" [Murre *et al.*, 2006, p. 244].

Clearly implicating the recurrent processes, Colzato *et al.* [2006] and later, Hommel and Colzato [2009], suggested that the initial process of combining co-occurring features is ad hoc binding resulting in 'object files'. Ad hoc binding is assumed to be a function of short term

memory. The second level of binding is to segregate and register the object files, which results in 'event files'. Hommel and Colzato [2009] held that two kinds of 'priming processes' control the memory for bindings. Offline priming happens before the stimulus occurs, due to factors such as mental set. Online priming is due to stimuli in LTM, such as familiar objects. All stimuli are detected in a non-selective feed-forward sweep. In the second stage, recurrent processes refine the input according to the principles governing the attention devoted to the task.

Based on the idea of reentrant mechanisms, Jaswal [2010, 2012] proposed that binding is a process of continuous refinement of a mental representation which is time consuming as well as attention demanding. This refinement entails consolidation of relevant features and a concomitant inhibition of irrelevant features, with relevance dictated by the contents of working memory and implemented physiologically by the reentrant processes in the brain. Depending on the features involved in binding, different kinds of bindings are envisaged, also requiring different kinds of attention [Jaswal and Logie, 2013].

Interestingly and importantly, this picture of binding by fast forward processing either by specialized set of neurons, or synchronous activity, aided by reentrant connections, more or less adequately explains binding in perception as well as memory.

2.2 BINDING AND SIMULTANEOUS PRESENTATION

Feature binding has been intensely researched since the 1980s, when the concept first evolved. Initially, several researchers [e.g., Duncan, 1980; Treisman and Gelade, 1980; Treisman and Schmidt, 1982] used simultaneous presentation in their studies due to different reasons. The prototype experiment for describing visual feature binding was "conjunction search" and that could perhaps be best explained with simultaneous presentation. Another reason for using the simultaneous mode of presentation was that real life situations apparently involve simultaneous presentation of a plethora of stimuli at all times. Lastly, binding is bifurcated into two types of problems: combination and segregation. Segregation can be best assessed and understood with simultaneous presentation.

Simultaneous presentation has the merit of possessing spatial information as well. Each stimulus can be encoded not only with respect to the absolute location it possesses in the visual scene, but also in relation to several other stimuli. Blalock and Clegg [2010] have shown the importance of configuration, which helps the whole scene to be encoded in visual short term memory. Jaswal and Logie [2011] have shown that change detection is easier when objects remain at the same locations from study to test due to configural encoding of these objects. The importance of locations in feature binding is discussed in a separate section ahead. This paragraph is just to remark and remember that location information is inextricably confounded with simultaneous presentation.

Another important consequence of simultaneous presentation is iconic memory. Although, a conjunction search paradigm deals purely with the encoding process and iconic memory has no role to play in it, in a change detection task, both encoding and maintenance processes are important and hence iconic memory becomes relevant. Sperling [1960] held that iconic memory is information persistence after the visibility of the stimulus is gone. Coltheart [1980] reviewed the mechanism of visible persistence and suggested that the visual stimulus evokes neural persistence, visible persistence, and information persistence, all three of which are rather distinct from each other. Typically, iconic memory lasts for about 100-300 ms [Loftus *et al.*, 1985; Loftus *et al.*, 1992]. Therefore, simultaneous presentation generates an iconic representation with spatial information, and this helps in detecting changes in binding during the change-detection task if the locations of study and test display remain the same. Nevertheless, if the locations change from study to test, then iconic memory may actually hamper performance. Clearly then, the effects of iconic memory are related to whether location

is a relevant factor or not. The effects of iconic memory can be obliterated by a mask presented immediately after the study display, although if the representation is already in visual short term memory, it is immune to the effects of such masks [Phillips, 1974; Sligte *et al.*, 2008; Smithson and Mollon, 2006].

A third factor that might enhance binding due to simultaneous presentation is attention. Desimone and Duncan [1995] reviewed evidence from their own lab in an influential article on mechanisms of selective attention and stated that competition arises when two objects fall in the common receptive fields of neurons. Due to this, objects suppress the distinctive neuronal response for each other. Attention biases this competition, as a result of which, the target object is enhanced and the distracters are suppressed. Thus, the biased competition model [Desimone and Duncan, 1995; Duncan, 1980, 2006] proposes that competition among stimuli is always good.

Nevertheless, the negative aspect of simultaneous presentation is that multiple stimuli compete with each other for a place in the limited capacity memory system. Emrich and Ferber [2012] suggested that binding errors [location and color] increase when objects are closer in space in simultaneous presentation. Contrasting simultaneous presentation of eight objects and sequential presentation of four objects in two consecutive displays, they found precision of memory binding was more with sequential presentation. Sequential presentation presumably prevents competition and removes interference from other features as well as objects.

Ahmad *et al.* [2017] question the idea that competition exists only in simultaneous presentation. They found that spatial proximity increases precision errors and misbindings, not only when the stimuli were presented simultaneously, but also when they were presented sequentially [when the second stimulus immediately followed the first]. However, the insertion of a 500 ms gap between the two stimuli eradicated the effect of competition. These researchers conclude that competitive interactions occur in simultaneous as well as sequential conditions.

In conclusion, simultaneous presentation provides a powerful cue in the form of spatial information. On the other hand, it also increases the chance of competition among stimuli and misbinding of features. The competition itself is presumably contingent on the spatial relationships among the simultaneously presented stimuli, besides recruiting biased attention. Therefore, it is an interesting research question to disentangle the effect of locations and simultaneous presentation in feature binding, behaviorally as well as at the neural level.

2.3 BINDING AND SEQUENTIAL PRESENTATION

Several converging lines of evidence [e.g., Brockmole *et al.,* 2002; Eriksen and Collins, 1967; Irwin, 1991; Phillips, 1974] show that stimuli presented in a sequence can be combined to form a single representation and there is an optimum interval between successive arrays, which promotes this integration. Eriksen and Collins [1967] presented their participants two apparently random dot patterns. However, if the dot patterns were superimposed, the resulting perception was of a nonsense syllable. The experimenters presented the dot patterns for 6 ms each, with the inter-stimulus interval [ISI] ranging from 25 to 100 ms, to find that the identification of the nonsense syllable increased as the ISI decreased. Apparently, the sensory effects of each stimulus presentation were summed up, or integrated visually, to reveal the nonsense word. They suggested that this integration is perceptual in nature.

However, Phillips [1974] showed the differences between iconic storage and VSTM, by presenting two visual checkerboard patterns, one after the other, to be judged 'same' or 'different' by participants, who were required to detect the extra cell which was lit up whenever there was a difference in the two patterns. He inferred the difference between initial, spatiotopic representations, which could also be masked; and later, VSTM representations, affected by set

size, but immune to masking. Phillips and Christie [1977], studying serial position effects in memory for sequences of matrix patterns, demonstrated the single item recency effect, which could be destroyed if the last matrix was followed by mental arithmetic, but was immune to masks. They concluded that the matrix patterns were stored in visual short term memory.

In the 1990s, Irwin [1991] used the 'empty cell detection task' first enunciated by Di Lollo [1980]. In this task, participants are shown a 4×4 grid comprising 16 cells. Seven of them are filled for about 33 ms, and after a blank interval, eight of the remaining cells are filled for 33 ms. The participant has to detect and identify the one cell, which remains unfilled in both displays. Using such a task, Irwin [1991] inferred that "trans saccadic memory" which allows one to do this task efficiently, is similar to the VSTM proposed by Phillips [1974] as it is not influenced by slight displacements of the grid, is not detailed, is limited in capacity, and is longer lasting [than the initial representation]. In another experiment, Brockmole et al. [2002] used delays of 0, 100, 500, 1000, 1500, 2000, 2500, or 3000 ms, between the two grids, and found optimum performance around 1500 ms. Beyond the delay of 100 ms, performance gradually improved, and leveled off between 1000 and 1500 ms. These studies showed that integration was happening in the VSTM.

Jiang and her associates, however, consistently argued against integration in VSTM. Jiang and Kumar [2004] used a variant of the empty cell recognition task to test whether there was one integrated representation in VSTM or there were two separate representations. They presented participants with two different arrays of dots, occupying two different sets of 6 locations in a 5×5 matrix. For the test, they queried participants whether a dot [marked by a small red square] was in a filled or empty location with respect to the previous two arrays. This probe was presented in three different contexts – either containing an integration of the two displays, or only array 1 or only array 2. Participants were neither encouraged nor stopped from integrating the two arrays. Results showed integration of the two arrays with an ISI of 0 ms, impaired consolidation of array 1 with an ISI of 200 ms, and separate representations of the two arrays with an ISI of 500 ms [and also at 1500 ms in a follow-up experiment]. Thus, they concluded that VSTM contains separate representations of arrays presented in a sequence. Jiang et al. [2005] also argued against integration and proposed that a process, which maintains the separate visual representations and compares them, is a better account of performance in the empty cell task. Kumar and Jiang [2005] reported that performance for the trailing array was better than that for the first array irrespective of whether memory for locations, colors, orientations, or conjunctions of color and orientation, was tested. The capacity estimate of VSTM remained about the same for these different types of stimuli.

A recent study using CDA by Li et al. [2020] reported that two sequential arrays were stored in different neural states if the interval between them was long [800 ms], but they were concurrently stored in the 'active state' at a shorter interval [400 ms]. But, this happened only if participants were doing the task for the first time. Once they form a mindset, they applied the same storage mode at all intervals.

All the aforementioned studies in this section are relevant in the present research because they show that successive stimuli can be integrated to form a single representation of a visual pattern only if the inter-stimulus interval between the two displays is not very large. That the participants might use a 'store and compare' process, is also important to understand performance in the sequential condition. The fact that there exists an optimum interval for integration is also relevant to the design of this research with sequential presentation. Nevertheless, the integration studies, do not specifically test feature binding in sequences. Integrating two arrays is quite different from binding the features of an object.

In the new millennium, several investigators [e.g., Allen et al., 2006; Brown and Brockmole, 2010; Brown et al., 2017] started studying feature binding with stimuli presented sequentially in a change detection task. Often they compared binding between simultaneously

and sequentially presented arrays. One common problem that all these researchers faced was how to control the exposure duration of stimuli presented simultaneously and sequentially. If the duration of exposure of both arrays is made equal, the time of the presentation of each item in the sequential array is necessarily reduced. But, if the exposure time per item is made equal to the simultaneous presentation condition, then the overall time for the sequential presentation condition increases. Researchers have tried to solve this problem in ingenious ways as will be clear from the review that follows. Most [but not all] of these studies provide evidence showing that performance is better in simultaneously presented arrays.

Allen *et al.* [2006] carried out an experiment in which they compared the accuracy of binding with simultaneous and sequential presentation. Participants were significantly better with simultaneous presentation. Since, Allen *et al.* [2006] presented each item at center screen, the next object presumably destroyed the representation of its predecessor. In consonance with this idea, serial position analyses done for the sequential condition showed the recency effect [change in the last item of the array had the highest hit rate]. This suggested that binding occurred with sequential presentation, but it was fragile, and simply decayed over time or subsequent stimuli easily overwrote the previous ones.

Other studies have also shown an advantage of simultaneous presentation over sequential presentation in feature binding. Brown and Brockmole [2010] compared memory for color-shape binding and single features, in older and younger people, with simultaneous and sequential presentation. The results were that performance was worse with sequential presentation, particularly for binding. Brown *et al.* [2017] compared younger and older people for single feature memory and binding using the simultaneous and sequential modes. They used the change detection task, and their stimuli were three colored shapes, but with a single probe presented at the center of the screen. They found better results for simultaneous than sequential presentation for binding as compared to single feature conditions, but only for younger adults in the age range 18 to 25.

Using a different task, which measures the precision of the binding response, Gorgoraptis *et al.* [2011] performed a series of experiments in which they compared simultaneous and sequential presentation at different locations, using a color-orientation binding task. In the memory array, 1 to 5 colored bars were shown for 500 ms, and then a blank screen for 500 ms. The participants had to manipulate the orientation of the probe, i.e., a single color bar, to match the orientation of the stimulus in the memory array. Findings indicated that the memory precision for sequential arrangements was lower, and the participants were prone to misbinding in this condition.

To the best of the present researcher's knowledge, the only study clearly showing superior memory for sequentially presented objects is by Yamamoto and Shelton [2009]. They used real life scenarios and found that the sequential presentation of objects makes it easy to memorize them. They used a room layout and six different objects. Participants were shown these objects either simultaneously for 30 secs, or sequentially for 2.5 secs per object, with the whole array being shown twice to equate the presentation time with the simultaneous condition. Results showed better performance with sequential presentation. Arguably, familiarity with these objects may have made it easier to recall them, even when they were presented in a sequence. Also, the repetition of the array may have been a factor in the better performance with sequential presentation.

Some studies have also shown a lack of difference between the two modes of presentation. In an early study, Shiffrin *et al.* [1973] compared simultaneous and sequential presentation at four different locations in visual field. They formulated the hypothesis that during sequential presentation attention is not shared among four locations, therefore it should

lead to better retention. However, results showed no difference between simultaneous and sequential performance.

Several studies [e.g., Ihssen *et al.*, 2010, 2014; Rudkin *et al.*, 2007], albeit testing memory for single features, have shown that the effects of sequential presentation can be made equal to those of simultaneous presentation. This can be achieved in many ways, such are devoting more resources to sequentially presented stimuli, reducing set size, etc.

Ihssen *et al.* [2010] tried to improve VSTM performance for uni-feature stimuli by presenting the sequential array in halves or repeatedly. In their experiments, they used two types of stimuli in arrays: 4 colored squares and 4 different shapes. They used three modes of presentation: simultaneous [e.g., set of four shapes above and a set of four colors below fixation], half-sequential [i.e., 4 items in one display], and the whole display repeated. Half of the trials were probed by the shapes and remaining half were probed by colored squares in the same locations. Finding demonstrated that the visual short-term memory was significantly improved by showing the stimuli in sequential or repeated manner. Later, Ihssen *et al.* [2014] carried out virtually the same experiment in fMRI. This fMRI investigation revealed that that the half-sequence or repeated mode increased brain activation in extra striate and primary visual areas relative to simultaneous presentation. This is understandable as the repeated presentation of the arrays, simply sends the same signal all over again to these primary visual areas.

The dual task paradigm has also been used to study the role of WM resources in the two modes of presentation. Rudkin *et al.* [2007] used the Matrix Patterns and Corsi Block tests as representative tasks of simultaneous and sequential presentation respectively to compare the two modes of presentation. They used the dual task procedure, using tasks highly demanding of central executive resources – random digit generation and random interval repetition [a vigilance task] – as secondary tasks in separate experiments. Results showed greater involvement of executive resources in the performance of the participants in the sequential rather than simultaneous mode of presentation. Fougnie and Marois [2009] used a change detection task in which participants had to detect the sameness of color, shape, either color or shape, and binding. During the retention interval, the participants performed a multiple object tracking task. Results found significant interference due to the secondary task in the binding condition with simultaneous presentation only. The impairment introduced by the secondary task was much less when the objects were shown sequentially at the centre of the screen. These studies show conflicting results regarding secondary task interference with simultaneous and sequential presentation. But, taken together, their results can also be interpreted in a different way, to suggest that the secondary tasks, drawing on central executive resource of attention, interfere more with performance in the condition which has a greater reliance on spatial memory in both experiments [Corsi blocks in the Rudkin *et al.* [2007] experiment and the simultaneous condition in Fougnie and Marois [2009] experiment]. Further, since locations are generally important for binding, it is understandable that greater disruption occurs in memory for bindings rather than for single features.

In another dual task study, Darling *et al.* [2009] used the letter P task, in which the letter P is presented within one of the various small square frames at several locations on the screen. They compared simultaneous and sequential presentation, testing two types of memory [location or font], with two interfering tasks – dynamic visual noise and tapping. As expected, the tapping task interfered more with memory for locations, whereas dynamic visual noise disrupted memory for the font of the letter. But the main effect of presentation was not significant, as it was just a single feature recall task. This type of result indicates that single feature recall [even if the feature is locations], is not influenced by a difference between presentation modes.

Thus, the overwhelming evidence that performance is better with simultaneous presentation rather than sequential presentation seems to come mostly from feature binding rather than single feature studies. Hoffman [1978, 1979] suggested that sequential information is directly/primarily represented in visual short term memory. He proposed a two-stage model for visual search. He suggested that perceiving a stimulus in the visual field takes place in two stages. In the first stage, multiple stimuli are extracted from background, and in the second stage, they are compared. In this second stage, all the stimuli are processed one by one, with focused attention. He emphasized results that show that slower rates of presenting a sequence cause rapid transfer from iconic memory to VSTM.

In complete contrast to this idea, Frick [1985] argued that VSTM stores simultaneous presentation but not sequential presentation. He compared both the modes of presentation with the task being verbal recall of consonants with and without articulatory suppression [speaking numbers 1, 2, 3, ... and so on]. He found that articulatory suppression caused confusion/errors only with simultaneous presentation. He concluded that visual short-term memory 'naturally' retains simultaneous presentations but does not 'naturally' retain sequential presentations. This idea was probably the result of the very common notion since early research that these presentation modes are differentially suited to different modalities. In an early experiment, Tulving and Lindsay [1967] compared simultaneous and sequential presentation for letters. They performed the experiment with both visual and auditory stimuli. They hypothesized that simultaneous presentation of multiple stimuli is also processed serially. They found higher performance in sequential presentation with both the modalities, but no evidence of sequential processing of simultaneously presented stimuli was reported. Most experiments in the 60s with sequential presentation used the auditory modality. While Eriksen and Spencer [1969] introduced the comparison of simultaneous and sequential modes of presentation as a paradigm in visual information processing, Shulman [1971] discussed the importance of comparing simultaneous and sequential presentation from the short-term memory perspective.

Penney [1975] reviewed modality effects for short-term verbal memory and suggested that sequential presentation has stronger associations with auditory memory. The major difference was in recent recalls. Recall rates decrease as the presentation rate increases in audition. However, when it comes to visual stimuli, the short-term memory buffer comes into the picture, retaining the visual sensations, and causing confusion. Krumhansl [1977] performed an experiment for detection of identity and location of two letters by manipulating modes of presentation and confusability. Accuracy was higher with simultaneous presentation and for the last letter in sequential presentation. This implied that spatial separation is more effective in simultaneous presentation than temporal separation in sequential presentation.

As mentioned earlier, Frick [1985] argued that the VSTM stores information for simultaneous presentation but not for sequential presentation. Kubovy [1988] also suggested that simultaneous presentation is efficiently perceived in vision while sequential presentation is better perceived in audition. Conway and Christiansen [2009] tested visual-spatial, visual-sequential, and auditory-sequential input, with moderate and fast presentation rates. They concluded that learning efficiency was better with visual-spatial and auditory-sequential, than visual-sequential input. Also, faster presentation rate only affected performance in the visual-sequential condition. Results suggested that spatial separation is better for discrimination in simultaneous presentation, and temporal separation is more effective for auditory sequential presentation. In sum, it appears that the visual scratchpad is for spatial but not temporal information. Further, these researchers insist that visual short term memory capacity is limited to a single pattern

However, Delogu *et al.* [2012] compared the memory for spatial locations and serial positions in the visual as well as auditory modalities, using the dual task paradigm. They found that accuracy in serial order recall was reduced by encoding item locations at the same time,

whereas the recall of item locations was unaffected by the concurrent encoding of the serial order. This asymmetric influence was similar for the auditory and visual modalities. This indicates that item location is processed relatively automatically, whereas maintaining serial order is more demanding of attention. The similar results for auditory and visual modalities suggest that binding serial order and location in working memory may involve common mechanisms across modalities.

Thus, it is debatable to what extent VSTM as merely a 'store' is used in sequential presentation conditions, and to what extent VWM processes and resources are used, especially to bind features. It is quite possible that very different factors and processes at the behavioral as well as the brain level explain binding with sequential presentation. In this context, it is interesting to note the results emanating from Harrison and Bays [2018]. They found that the ability to maintain visual features in memory with sequential presentation is unaffected by cortical spacing in the visual cortex, suggesting a role for higher cortical areas such as posterior parietal or prefrontal regions or as yet unspecified mechanisms, according to which stimulus features are bound to their temporal order. Perhaps the present experiments, and the neural basis of simultaneous and sequential presentation, will throw some light on these questions.

2.4 THE IMPORTANCE OF LOCATIONS

Spatial location has always been considered important for visual processing. The feature integration theory [Treisman and Gelade, 1980, Treisman and Schmidt, 1982] ascribed a special role to locations in conjunction search or feature binding. Treisman and Sato [1990] proposed a 'spatial map' that enables whatever is present in any location to be bound together with the glue of attention. Sagi and Julesz [1985] held that locations or other features can be detected in parallel but identifying features and 'knowing' them requires serial, focal attention. Wolfe [1994] revised the guided search model to give a special place to locations as a feature, quite distinct from other features. Initially, Wolfe *et al.* [1989] had proposed that parallel processing of multiple features guides visual search. The more the number of features in a target, more would be the processes working in parallel. Further, the number of parallel processes was proposed to be directly proportional to the efficiency of search. The revised model [Wolfe, 1994] held that other features are processed in parallel, but at any given time, a few locations are processed together, thus making processing of locations different from other features.

Several behavioral studies demonstrate the special nature of locations as compared to other features. [Ashby *et al.*, 1996; Treisman and Schmidt, 1982]. Johnston and Pashler [1990] slightly modified the visual search experiments of Treisman and Gelade [1980] and proposed that binding calls upon the attention signal only for locations. Cohen and Ivry [1991] experimented on the factor of relative locations in binding and found that recognition is relatively faster for objects that are apart from each other. Wheeler and Treisman [2002] found that participants have better memory for locations than for color or shape. Keele *et al.* [1988] also pointed out that temporal synchrony is not sufficient for binding and that binding cannot take place until all the features belong to the same location. Behavioral studies showing the influence of iconic memory in retention of bindings also show the importance of spatial configuration [Jaswal, 2010, Jiang, Olson, *et al.*, 2000; Phillips, 1974]. Logie *et al.* [2011] showed that randomizing locations was more disruptive of memory for bindings as compared to randomizing colors or randomizing shapes. In contrast to these studies which emphasize obligatory encoding of stimuli in terms of their locations, Udale *et al.* [2017; 2018a] insist that objects are not automatically coded in relation to the background or other irrelevant objects, and in fact, retrieval may be strategy driven [Udale *et al.*, 2018b].

At the physiological level, Milner and Goodale [1995, 2008] proposed that there are two different pathways in the brain for processing visual input. The dorsal pathway processes the

visual input from V1 occipital region to parietal cortex via the lateral geniculate nucleus [LGN], visual area 3 [V3], and middle temporal visual area [MT]. This pathway processes features like spatial location and motion. The second pathway is the ventral pathway, which goes from visual area 1 [V1] to the temporal cortex via visual area 2 [V2], visual area 4 [V4], and the inferior temporal cortex [IT]. This pathway processes the surface/static features of an object such as color, shape etc. Vidyasagar [1999] also proposed a model in which the magnocellular [dorsal] pathway contributes to the parvocellular [ventral] pathway, and which causes binding to happen.

Schneegans and Bays [2017] proposed a neural model for feature binding in visual working memory based on several populations of neurons responding to conjunctions. Their model and experimental results yield the conclusion that non-spatial features are also bound only through their shared locations. There was no evidence of direct binding between color and orientation in their experiments. Thus, they also argue for and provide evidence for a special role of location in feature binding, and their model explains how this special role works in the neural architecture. Recently, Fang *et al.* [2019] demonstrated that in the retinotopic phase, when a location was being attended, the adjacent locations were suppressed. This is known as surround suppression, and this sensory mechanism may also contribute in making location special. More important, however, is the role of the brain regions coding the environment according to a spatial map. Grid cells in the entorhinal cortex and place cells in the hippocampal region are the specialized neuronal networks that encode a spatial map [Jacobs *et al.*, 2013].

The processing of location information is also different from other features in involving the parietal cortex. Damage to the parietal cortex can cause decreased performance in change detection tasks involving locations [Pisella *et al.*, 2004; Robertson *et al.*, 1997; Treisman, 1998]. Parietal contribution is very well established for spatial attention [e.g., Colby and Goldberg, 1999; Posner and Peterson, 1990]. Shafritz *et al.* [2002] confirmed the role of the parietal cortex during binding with simultaneously presented stimuli. Since the parietal cortex is associated with processing of locations and it is shown to be active in feature binding by several studies, as per this line of evidence also, one can conclude that locations are especially important in feature binding. The question of interest regarding locations in the present research is how far they are involved in, or modulate, the effects of simultaneous and sequential modes of presentation in feature binding.

2.5 THE ROLE OF ATTENTION IN BINDING

Several researchers have debated whether the binding process is non-conscious and automatic, or a conscious, resource demanding process. Electrophysiological as well as neuroimaging evidences support the idea that binding is a resource demanding process [Jing *et al.*, 2011]. Scientists have found the neurotransmitter acetylcholine to be responsible for binding through experiments on rodents [Botly and Rosa, 2008, 2012]. One of the major functions supported by acetylcholine is decision-making using spatial attention [Angela and Dayan, 2005]. Such evidence indicates the involvement of spatial attention in feature binding.

It must be emphasized that it is not the arousal function of attention that is important for binding. Rather, it is the selective property of attention, which is important. In fact, high arousal may be detrimental to binding. Mather *et al.* [2006] tested binding of pictures and locations. The pictures were classified in three categories: low, medium, and high arousal. They found that the accuracy of binding was inversely proportional to arousal. The study was done in the fMRI environment and they reported higher activity in the fusiform gyrus, middle temporal gyrus, middle occipital gyrus, and the lingual gyrus, with medium and high arousal conditions, suggesting that more arousing stimuli are associated with dominant activation in these occipital, early processing areas. Note that feature binding is usually associated with the parietal areas.

The feature integration theory [Treisman and Gelade, 1980] clearly stated that spatial attention acts as 'glue' to bind all other features. Recent studies such as that by Dowd and Golomb [2018] show that binding survives with the power of spatial attention, even when the focus of attention is dynamically shifting in the visual field. Harris *et al.* [2019] also highlighted the importance of spatial attention and suggested that it can protect the organism from misbinding of features.

Duncan [1980] asserted that only targets [defined by the templates in WM] are selected for further processing by attention, whereas all other objects which are not targets are identified as such and rejected very early on by parallel, unconscious processes. Desimone and Duncan [1995] proposed the biased competition theory according to which WM provides a template according to which relevant sensory input is selected for processing. The assumption is that top down attention biases the competition among otherwise similar objects in the environment. Based on studies of monkeys using recordings from single units [Chelazzi *et al.*, 1993], it was proposed that sensory input at any point of time manifests incessant competition. The representations in WM bias this competition such that inputs matching the 'templates' in WM are the ones which are processed further. Duncan [2006] held that competitive bias can begin anywhere in the system and spread to higher and/or lower levels, but he also reiterated the important and directive role of a pattern of fronto-parietal activity called the "multiple demand pattern" as it results from many different kinds of cognitive demands. His theory is often held to be the archetype of top-down processing ideas.

Nevertheless, while attention selects the relevant objects for further processing and denies representation to the irrelevant objects, the *features* of the relevant objects are processed in an all or none fashion. Duncan [1996, 1998, 2006] held that features are encoded across different cortical regions, albeit in an integrated way. Therefore, if attention is directed to one feature, all features of the object, whether relevant or irrelevant, become dominant in their particular regions of the brain. This idea was supported by fMRI data by O'Craven *et al.* [1999a, 1999b] who found activation in response to the attended as well as non-attended task irrelevant attributes of a stimulus in their studies. More clear support came from a study by Schoenfeld *et al.* [2003] who combined recordings of event related potentials and fMRI. They found that even the irrelevant features activated quite early, and participated in the perceptual integration of all features of the attended object. Using event related potentials, Winkler *et al.* [2005] reported evidence that relevant as well as irrelevant features are integrated without attention "normally" in visual as well as auditory modalities. Nevertheless, attention is indispensible for correct binding under exceptional conditions, such as when many stimuli, or complex stimuli, are processed in relatively short time frames. Recently He *et al.* [2020] also provided evidence that a secondary task consuming object-based attention led to a larger impairment in bindings relative to constituent single features. Such evidences indicate the automatic nature of visual feature binding, in which all features are obligatorily encoded in the initial representation of the object.

The argument that competition is necessary for attention to develop, is supported by neuroimaging evidence that the posterior parietal cortex, associated with focused visuospatial attention, promotes feature binding only when there is simultaneous presence of other objects. Kastner *et al.* [1998] used fMRI evidence to substantiate that when stimuli are simultaneously presented, their cortical representations interact in a competitive and suppressive way in the ventral [object recognition] pathway. However, this was not evident when stimuli were presented sequentially. In a second experiment, spatial attention focused on the objects was found to counteract the suppressive effect, and more so in the simultaneous as compared to the sequential presentation condition. Using fMRI, Shafritz *et al.* [2002] established that the posterior parietal cortex was active when multiple objects were simultaneously presented, but not when they were sequentially presented in the same location at the center of the screen.

It is clear that attention has a great role to play when multiple stimuli are spatially segregated. But, the role of attention in temporal separation of stimuli is still not clear. When stimuli are presented one by one, the absence of distracters reduces the importance of spatial attention, but attention is presumably still required to bind that object. Robertson [2003] reviewed the neurological and behavioral evidences for binding and stated that binding across space definitely needs the attentional component from the parietal lobe. Binding across time does not really evoke the same responses in the parietal lobe, although it also needs the spotlight of attention. Therefore, there are additional neural components, which are involved in the process of binding of sequences, possibly from the ventral stream. Evidence regarding 'temporal attention' was reviewed by Chun *et al.* [2011], who suggested that it has similar properties to spatial attention. Just like the limited capacity of spatial attention, temporal attention also functions by selecting task-relevant information, but it does so over time. Rapid serial visual presentation is a good way to study temporal attention. The reports in literature collectively show that in any serial presentation, attention is required in order to get that stimulus in visual memory. Thus, attention should not be understood as uniquely dependent on spatial configurations.

The role of attention in feature binding in working memory has also been extensively studied by Baddeley and colleagues. In the original working memory model [Baddeley and Hitch, 1974] the central executive provided the attentional control required for processing information in working memory. Attention was also held to be necessary, and indeed, indispensible for the 'episodic buffer', proposed by Baddeley [2000]. It was the new component of WM, which combined information from different modalities, and where feature binding also took place. Baddeley *et al.* [2011] reviewed the idea of the episodic buffer by scrutinizing the impact of a range of concurrent tasks on binding and single feature memory. There was simply no differential effect of these tasks on bindings in comparison to single feature tasks. Nevertheless, bindings were more vulnerable to disruption by a subsequent stimulus [a suffix]. Ueno *et al.* [2011] found that a plausible suffix to the study display had a greater effect than an implausible suffix [from outside the memory set], although whether the suffix itself had a single feature or both features from the memory set, did not differentially affect performance. Thus, they concluded that an early feature based filtering process in WM was followed by an object-based storage, where earlier items were completely displaced by later ones. To test the idea that feature binding requires additional WM resources in comparison to memory for single features, Allen *et al.* [2012] also studied the effect of different kinds of load, i.e., articulatory suppression and backward counting. Findings clearly showed that feature binding did not require any additional resources, leading the authors to infer that binding probably occurred outside the episodic buffer. Allen *et al.* [2014] emphasized a two stage processing of bindings. In the first phase, every object gets the benefit of attention, and in the second phase, the maintenance of the current and subsequent object demands executive resources. They also suggested that in a sequence, if executive attention is directed towards another task or to a subsequent stimulus, then the previous stimuli is likely to disintegrate.

Hitch *et al.* [2020] recently reviewed their own work and suggested that poor retention of the older stimulus in a sequence may be a result of limited executive resources. By this account, the binding of different features is an early stage process. After the bound object enters visual short term memory, the maintenance phase starts. Thus, the 'problem' of binding is divided in two phases, one is binding when encoding takes place, usually tested in conjunction search tasks, and the other is maintenance of the encoded bindings, usually tested with change detection tasks. The feature integration theory elaborated the role of attention in the initial phase of binding of features. It explains binding in the encoding phase, but is does not really address the maintenance phase of binding.

Such ideas accord well with the contention of Luck and Vogel [1997] that retention of multi-feature object in memory requires no extra resource than maintaining a uni-feature object. Other researchers also held that maintenance of bindings in visual working memory is possible without attention [e.g., Delvenne *et al.*, 2009; Eckstein, 1998; Gajewski and Btockmole, 2006; Mordkoff and Halterman, 2008].

Encoding and maintenance phases in sequential presentation condition are also distinguished by Ricker and Cowan [2014]. They argued that proper consolidation of the memory array is needed to help tighten the bond of features. In their experiment on memory for single feature objects, improved memory performance was obtained with sequential presentation, by introducing a blank interval after each stimulus presented in the sequence.

Feature bindings are even more resource demanding than single features, as shown by Fougnie and Marois [2009]. Evidence consonant with this has been given most recently by He *et al.* [2020] using color shape binding of sequentially presented stimuli as the primary task and object based tasks as secondary tasks introduced in the maintenance phase. In one experiment, this was a transparent motion task and in the other experiment, it was a mental rotation task. Both were disruptive of binding performance more than single feature memory. In fact, single feature memory was virtually immune to these secondary task manipulations.

Several studies which have used sequential mode of presentation of stimuli have also analyzed the serial position effects. These researchers usually find a recency effect, which is the superior retention of the last item in a sequence [e.g., Allen *et al.*, 2006; Atkinson *et al.*, 2018, 2019; Berry *et al.*, 2018; He *et al.*, 2020; Hitch *et al.*, 2018; Hu *et al.*, 2014; Hu *et al.*, 2016; Maxcey-Richard and Hollingworth, 2013; Phillips and Christie, 1977]. This may be because the most recent stimulus is in a highly accessible privileged state such that it can be visualized and thus retained in a relatively automatic and resource-free manner, whereas the other stimuli are in a non-privileged state, and can be maintained in memory if required, only by devoting extra attentional resources.

Relevant to the automaticity vs. resource demanding debate is also the evidence related to learning of bindings so that they are represented in the permanent memory store. Logie *et al.* [2009] and Shimi and Logie [2019] showed that the Hebb repetition paradigm did not yield any improvement in a change detection task testing color-shape-location bindings, but cued recall and reconstruction did lead to some improvement over trials. They argued for 'fragile' memory representations of binding, which nevertheless, generate a weak episodic trace that becomes stronger over trials and is presumably aided by awareness.

Also, whereas memory for feature bindings may not be stable from trial to trial, there is some evidence that practice does influence search for bindings. Sireteanu and Rettenbach [2000] stated that searching for a conjunction becomes efficient after prolonged practice. Perceptual learning interacts with other factors such as selective attention, stimulus context, and stimulus familiarity and all these factors have an impact on visual feature binding. Recent research shows that stimulus discrimination is particularly dependent on perceptual learning. Not surprisingly, Xie and Yu [2020] suggested that learning with multiple stimuli is more efficient than learning with a single stimulus. This is one probable reason why perceptual learning helps in stimulus discrimination. Zhang and Li [2020] report empirical evidence for the fact that perceptual learning improves discrimination of orientation of objects. It is also relevant to note here that strategic influences on feature binding have also been demonstrated [Atkinson *et al.*, 2018; Udale *et al.*, 2018b].

Thus, there are ongoing debates on various aspects of the maintenance phase of binding, such as whether the objects are merely stored in VSTM or whether concurrent processing is also taking place. At least, for sequentially presented stimuli, both maintenance [of the older stimulus] and encoding [of the new stimulus] seemingly take place at the same time.

Differentiating between encoding and maintenance suggests that sequentially presented stimuli need extra attentional resources for binding. The long-term retention of bound items, despite fragile representation in short term memory also needs to be explained.

Perhaps, these factors can be better understood by examining the binding process in temporally distributed items. The difference in resource usage during simultaneous and sequential presentation needs to be studied at the behavioral and neural levels to understand the contribution of attention in the binding process.

2.6 NEURAL CORRELATES OF BINDING

The present research was undertaken with the premise that two lines of evidence – behavioral as well as fMRI – will together yield important insights regarding the role of locations and mode of presentation in feature binding. In the foregoing review of literature, both these approaches were discussed together, often drawing substance from each other. Nevertheless, as one objective of this research is to demarcate the brain areas involved in feature binding with stimuli presented with simultaneous and sequential modes of presentation, it is important to specifically review the areas showing activation in feature binding, particularly with simultaneous and sequential presentation. This will also help in the delineation of regions of interest [ROI] in the fMRI experiment.

2.6.1 Neural correlates of visual feature binding

The exact neural co-relates of binding are still unclear despite visual feature binding being extensively investigated. Initially a stimulus is processed in the primary visual cortex. Zeki [1978] exposed the visual areas in the brain of primates and found that features such as shape, color, motion, etc., were processed in separate visual areas. Van Essen [1985] showed that there are multiple pathways from the retina to feature specific brain regions, such that perception of form follows a separate pathway from the perception of motion, and so on. Such evidence indicates that feature binding is very much a cortical rather than a sensory mechanism. Indeed, many would opine that the binding problem exists because features are processed in disparate areas of the brain. Milner and Goodale [1995; 2008] popularized the idea that visual processing takes place in two streams – dorsal and ventral – which go from the primary visual area, to the parietal, and the temporal regions, respectively. Not surprisingly, the parietal and temporal areas are both involved in visual feature binding.

The bilateral parietal regions are implicated in visual feature binding in multiple studies. [Ashbridge *et al.*, 1997; Bernstein and Robertson, 1998; Cohen and Rafal, 1991; Friedman-Hill *et al.*, 1995; Humphreys, 2001; Shafritz *et al.*, 2002]. Typically, these studies contrast performance in binding and single feature conditions and study patients with damage to the area under investigation, or study the underlying areas of the brain showing activation in fMRI. For example, Donner *et al.* [2002] compared the binding of color and orientation with easy feature [color] and hard feature [orientation], and concluded that the response of the frontal cortex and intra parietal sulcus was dominant in the binding condition. The involvement of fronto-parietal networks shows that the process is multi-step rather than instantaneous. Wei *et al.* [2011] compared conjunction search with single feature search and supported the involvement of the intra parietal sulcus in conjunction search. They tested three degrees of conjunction search, i.e., within dimension, cross dimension, and combination of within and cross dimension conjunction search. They found that the parietal region was active during all three conditions when contrasted with single feature search.

Another account of feature binding is the involvement of the temporal cortex and temporo-parietal junction [Baumgartner *et al.*, 2013; Pollman *et al.*, 2014]. Early studies showed that the MTL is necessary for maintaining information over long delays [Cermak, 1976; Chun

and Phelps, 1999; Giovanello *et al.*, 2003; Winocur and Weiskrantz, 1976]. But Stark and Squire [2001] suggested that MTL may be implicated in maintaining information over shorter durations also. Mitchell *et al.* [2000] compared feature binding in younger and older people and found greater activation in the anterior hippocampus on combination trials in younger people rather than older people. They tested the binding of a specific object and its location in a sequence of three displays and found that younger people are better in binding with greater activation in the hippocampus region in the brain while older people are not good at binding and show significantly lesser activation. Olson *et al.* [2006] confirmed that the MTL is necessary for conjunction of object features distributed temporally over one second as well as eight seconds. These researchers compared the accuracy of feature conjunction between patients with MTL amnesia and healthy controls. They performed a single feature recognition task and conjunction of object and location task with patients as well as controls. Findings show that patients performed equal to controls in single feature recognition but dramatically less in the conjunction task. Their results indicated that deficits in feature binding in working memory are because of damage to the hippocampus. This study also suggests a role for MTL in memory maintenance for eight seconds, although not for one second.

Thus, taken together, the majority of studies of binding highlight the involvement of areas from the parietal and temporal regions. The parietal regions are responsible for attention whereas the temporal regions are associated with memory. Both these cognitive processes [attention and memory] are integral to feature binding. The close connection between binding, attention, and memory is also illustrated by gamma band oscillations in electro encephalography [EEG] studies. Jing *et al.* [2011] found gamma band oscillations to be associated with the binding process. Earlier, gamma band oscillations have been found responsible for attention and memory [Tallon-Baudry *et al.*, 1998]. Therefore, both these regions are of specific interest in the present work.

2.6.2 Neural correlates of binding with simultaneous and with sequential presentation

In the context of the aim of the present research to contrast the effects of simultaneous and sequential presentation in feature binding, it is also important to note the common and distinct areas activated with simultaneous and sequential presentation of stimuli.

Among the seminal fMRI studies, Kastner *et al.* [1998] studied visual processing of complex stimuli presented either simultaneously for 250 ms followed by a blank interval for 750 ms or sequentially for 250 ms each. They targeted only the visual areas. Overall, activation was more with sequential presentation in the primary visual areas V1, V2, V4, and the temporo-occipital area. This was true when participants were asked to attend to the stimulus locations as well as when they were asked to attend to another task presented at fixation. Kastner and Ungerleider [2001] suggested that attention caused increment in activation with both the presentation modes, with areas V1 and V2 showing little or no increment, but areas V4 and the posterior part of the infero-temporal cortex [TEO] showing larger differences in favour of sequential presentation.

In research on feature binding, the differences between simultaneous and sequential presentation are particularly apparent in studies regarding the parietal cortex. Friedman-Hill *et al.* [1995] tested the ability to perform conjunction search in the patient R.M. who had bilateral parietal damage. The researchers tested color-shape binding with simultaneous and sequential presentation. In the sequential presentation condition, they presented the stimuli one by one in a fixed single spatial location for 500 ms each. Under these conditions, R.M. could report the target correctly. Thus, R.M. was more prone to making conjunction errors in simultaneous presentation but not in sequential presentation. This implies that R.M. could not use spatial information, but could easily use temporal coincidence in sequential presentation. Also, they did not report any enhanced activation during binding with sequential presentation.

Shafritz *et al.* [2002] investigated the brain regions involved in feature binding through fMRI. They presented two objects simultaneously at different locations and sequentially at the same location and tested brain regions involved in single feature detection and feature conjunction, i.e., binding. Interestingly, they found an enhanced activation in the right superior parietal cortex and anterior intra parietal cortex for the binding task than single feature detection, but only in the simultaneous mode of presentation.

These studies, which show the involvement of the parietal areas in binding with simultaneous presentation only, may do so because the parietal cortex is linked to locations as well as attention. Areas from the parietal cortex are well-known for spatial attention-related activity [evidence reviewed by Colby and Goldberg, 1999]. Xu and Chun [2006] demonstrated a role for the parietal regions not only in perception, but also memory, using fMRI. They found that only about four objects in different locations could be represented in the inferior intra-parietal sulcus. However, the superior IPS and the lateral occipital complex were variable in storage capacity. These areas maintained less than four objects as the complexity of the objects increased. These neural response patterns were observed during VSTM encoding as well as maintenance. Xu and Chun [2009] theorized that whereas the inferior IPS maintains spatial attention over a fixed number of objects at different spatial locations and is thus responsible for "object individuation", the superior IPS and the lateral occipital complex encode and maintain a variable subset of the attended objects, depending on their complexity, and are therefore responsible for "object identification". Kawasaki *et al.* [2008] also reported that the posterior parietal cortex maintains features such as color, shape and motion in VSTM. When all these features are superimposed [as in binding stimuli] memory was reduced. The dominance of the parietal contribution in feature binding investigations validates the role of spatial attention in the binding process. Indeed, the parietal cortex is specialized to process spatial information and attention. Binding of features is necessary for differentiating between the objects. May be, parsing ambiguous neural signals happens on the basis of their spatial relations.

Nevertheless, there is some evidence for parietal involvement when viewing sequential stimuli as well. Di Pellegrino *et al.* [1998] found evidence for extinction in a patient due to damage to the middle temporal gyrus as well as inferior parietal lobule. Arend *et al.* [2011] compared patients with right cortical lesions with healthy controls in temporal integration using an RSVP task. They presented a stream of five alphabets in random order from trial to trial. Participants were required to recognize the white alphabet among black distractors. They concluded that patients made more errors in right as well as left visual fields than controls. They observed that the right parietal cortex controlled spatial as well as temporal integration.

Besides the parietal cortex, the MTL is another area linked to processing of objects in a sequence. Pertzov *et al.* [2013] conducted two experiments, one with simultaneous presentation, and the other with sequential presentation. In the first experiment with simultaneously presented objects, they tested the binding of location with a fractal shape, the participant being asked to drag a fractal to its location. In the second experiment, the participant had to remember the binding of color and orientation of three sequentially presented bars in a memory display and adjust the orientation of the test probe of a particular color. They performed these tasks with voltage-gated potassium channel complex antibody patients [who have symptoms of amnesia, confusion, and seizures], and who showed damage to the MTL, and compared them with healthy controls. Patients performed equal to controls when memory arrays consisted of one bar but their performance dramatically decreased as the number of bars increased to three. They were also poorer in memory for bindings as compared to single features. These results indicate that MTL is necessary for binding of multiple objects, regardless of whether they are presented simultaneously or sequentially. It is clear that MTL is not only relevant to long-term memory, but plays an important role in binding of features separated over time even with short delays such as 500 ms.

Using a visual search paradigm, a direct comparison of simultaneous and sequential presentation was made by Coull *et al.* [2003], but using a search paradigm. They used fMRI on eight participants taking two independent variables, distribution of distracters, and type of search. Distribution of distracters was spatial [simultaneous presentation] or temporal [sequential presentation], whereas type of search was either feature search or conjunction search. Brain regions, which were activated during spatial conjunction search, were the right dorso-medial parietal cortex in the region of the precuneus, and the right ventro-medial extra striate cortex. Brain areas, which were activated in temporal conjunction search, were the left prefrontal cortex, frontal operculum, and putamen bilaterally, and right superior temporal cortex. Thus, in this study, sequential presentation as compared to simultaneous presentation recruited anterior brain areas such the frontal and temporal cortex linked to memory and other executive functions.

Jenson and Lisman [2005] in a review claimed that the process of binding during sequential presentation is associated with long term potentiation [LTP] in the hippocampus. Memory encoding is thought to occur by LTP in the hippocampus; LTP underlies learning of memory sequences even if they are far apart in time, if it is conceptualized that the input to the hippocampus is from a cortical multi-item working memory buffer in which theta and gamma oscillations have an important role. In such a buffer, memory items that occurred seconds apart are represented with a temporal separation of 20–30 ms, thereby bringing them within the LTP window of about 100 ms in the hippocampal cells.

To conclude this section, it can be inferred that the involvement of the parietal region is dominant in binding when objects are presented simultaneously. Simultaneously presented objects automatically evoke spatial attention and this causes significant activity in the parietal region. Whether and how far there is parietal involvement in sequential presentation will be observed in the present work. There is some evidence that besides the parietal areas, the temporal and frontal regions are involved in sequential presentation of conjunctions [Coull *et al.*, 2003]. Thus, this research will also assess the brain areas associated with sequential presentation. Above all, the aim is to study if locations being unchanged and random modulate the effect of simultaneous and sequential presentation in these brain areas. This interaction effect will be assessed in specific regions of interest [ROI] which earlier research has indicated to be important in feature binding. These ROIs, and the studies on which they are based, are described in the introduction to the fMRI experiment [Chapter 6].

2.7 THE CONFOUND IN LITERATURE AND OBJECTIVES OF THE PRESENT WORK

This research primarily aims to compare the effect of simultaneous and sequential presentation on feature binding, with locations being unchanged or random from study to test in a change detection task.

Many studies of feature binding have used the simultaneous mode of presentation. It is important to realize that simultaneous presentation allows the utilization of location as a powerful cue for feature binding. The importance of location in binding has been emphasized by the feature integration theory [Treisman and Gelade, 1980; Treisman and Sato, 1990] as well as the guided search model [Wolfe, 1994]. It has been shown that location is better remembered than colors [Wheeler and Treisman, 2002]. Studies also show that bindings are more vulnerable to location change and suggest that locations play a central role in early maintenance and retrieval of bound objects [Logie *et al.*, 2011; Treisman and Zhang, 2006]. For this and other reasons, simultaneous presentation of multiple objects was initially considered crucial for binding studies.

At the turn of this century, however, researchers started to contrast simultaneous and sequential modes of presentation in feature binding tasks. Allen *et al.* [2006] used a shape-color

binding task with both modes of presentation. Results showed that performance was less accurate with sequential mode of presentation. Others [Brown and Brockmole, 2010; Brown *et al.*, 2017; Gorgoraptis *et al.,* 2011] have also reported binding deficits with the sequential mode of presentation. Perhaps the only study showing better recall of objects with sequential presentation is by Yamamoto and Shelton [2009], who used real life scenarios.

Turning to neural evidence for the difference between sequential and simultaneous presentation, an early study by Shafritz *et al.* [2002] used fMRI to investigate the brain regions involved in the binding process with both presentation modes. They presented two unique shapes with different colors [out of a pool of five colors and shapes] in the study display. The two objects were presented sequentially as well as simultaneously. In simultaneous presentation, the two objects were shown above and below fixation for 180 ms. In sequential presentation, the objects were presented at center-screen for 140 ms with an inter stimulus interval of 300 ms. The results of fMRI showed that the right superior parietal cortex and the intra parietal cortex were more activated for binding than for single features, but with simultaneous presentation only, in which multiple stimuli were shown in different locations.

It is possible that simultaneous presentation generally yields better performance in experimental tasks, simply because it allows configural encoding. On the other hand, simultaneous presentation also increases competition, and the errors from within the memory set are more common than when stimuli are presented sequentially [Emrich and Ferber, 2012]. This might be a factor in better or equivalent performance with sequential presentation, especially in real life conditions, where experience or familiarity with stimuli might mitigate the effects of competition, and increase the distinctiveness of stimuli.

However, none of these studies manipulated locations independently of presentation mode. In some studies [e.g., Allen *et al.*, 2006] stimuli were presented at the same fixed locations for study and test. Yet other studies relied on a single probe at a fixed location, to control the influence of locations. In the experiment by Gorgoraptis *et al.* [2011] locations were random in the study display in each trial, in the sequential as well as simultaneous presentation modes, thus they were presumably irrelevant to the participant's memory for the orientation of the different colored bars, tested by a single probe. But, if the stimulus is encoded with locations included, this does not eradicate its influence. Also, in the design of this study, mode of presentation and location were not completely crossed. Notice that there was no experimental condition, which tested simultaneous presentation with same locations. Finally, none of these researchers endeavored to study the underlying brain areas and mechanisms through fMRI.

To unravel the effects of mode of presentation and relative locations, it is important to independently/orthogonally manipulate these two variables. Earlier, Jaswal and Logie [2011] studied simultaneous and sequential modes of presentation in separate experiments keeping locations constant in one condition and randomizing locations from study to test in the other condition. Performance was inferior with sequential presentation when the participants never saw all the stimuli together in the study display, even when locations of the stimuli remained the same. This suggests that simultaneous presentation results in better performance, because it gains from the relative location information, which is concomitant with simultaneous presentation. In fact, when locations were random, and thus rendered irrelevant to the task, there was no significant difference in performance between simultaneous and sequential presentation experiments. Nevertheless, simultaneous and sequential presentation modes were not directly compared in these experiments. Also, there was no attempt to use any physiological measure.

In conclusion, it is assumed that an important strategy for extricating the effects of mode of presentation and location is to manipulate both of them as separate independent variables. Thus, the main objective of the present research is to compare simultaneous and sequential

presentation when stimuli are shown in the same locations, and when they are presented in random locations, from the study to test display. This is done through behavioral experiments as well as in the fMRI environment.

It is expected that change detection performance will be better with unchanged locations than random locations, and with simultaneous presentation than sequential presentation. More importantly, an interaction effect is predicted such that the difference between unchanged and random locations will be greater for the simultaneous presentation condition than for the sequential presentation condition. The serial position effects observed in the sequential presentation condition are expected to add fine-grained detail regarding the use of the temporal code provided by sequential presentation in all experiments to the participants.

The fMRI experiment is expected to uncover the brain areas involved in four experimental conditions. Major questions of interest are whether or not sequential and simultaneous modes of presentation recruit the same areas of the brain, whether location and attention related areas are activated, and if so, are they recruited to a different extent in the different experimental conditions. With regard to the differences in the neural correlates of performance in the two modes of presentation, it is hypothesized that greater activation will be observed in sequential presentation as compared to simultaneous presentation, possibly involving the areas associated with memory for objects. It is also expected that when comparing random locations and unchanged locations, random locations will be more attention demanding, and consequently recruit more attention related areas. The interaction of mode of presentation and locations will be explored in specific regions of interest related to feature binding in previous studies.

The next chapter delineates and discusses the method used in the present research. The behavioral experiments are described Chapters 4 and 5, whereas the fMRI experiment is reported in Chapter 6. The results are discussed and then summarized in the two final chapters.