Is There Hebb Repetition Learning for Semantic Information?

Rebecca Legg
Doctoral College
Bournemouth University

Thesis submitted for the degree of Master of Research in Psychology

January 2023

Copyright Statement

Abstract

There are numerous pieces of empirical evidence (Hebb, 1961; Johnson et al., 2017; Page et al., 2013) showing that repeating a sequence of information results in better serial recall accuracy relative to non-repeated sequences (i.e., the Hebb repetition effect, HRE). Additional information, other than exact item and order repetition, can produce HREs such as motor responses (Johnson et al., 2017) and metrical patterns (Paice et al., in preparation). There has, however, been no investigation as to whether semantic information can be acquired in a similar way to item and order information, despite evidence showing semantic similarity improves recall in Immediate Serial Recall (ISR) tasks (Saint-Aubin et al., 2005). The current research therefore investigated whether a HRE for semantic information exists across three experiments using HRE procedures. Experiment 1 showed no evidence of a HRE when only a semantic pattern was repeated, (i.e., without exact item repetition). Experiment 2 replicated the canonical HRE with exact list repetitions; however, the recall advantage generated did not transfer to novel item lists following the same semantic pattern. That is, Experiment 2 showed that once learnt, sequence knowledge was not transferred to a semantically related list. Lastly, Experiment 3 adopted a typical Hebb repetition paradigm wherein participants learnt lists of category labels; however, at test participants reconstructed the lists using exemplars of the category labels. There was a significant HRE, however, the exact mechanism driving this effect is unclear. In general, results are discussed alongside two prominent models of the HRE (Burgess & Hitch, 1999, 2006; Page & Norris, 1998, 2009). Overall, findings across the three experiments suggest that semantic information is not acquired in a similar way to item and order information when learning new lists, therefore, supporting the two models in their current form as explanations of the HRE.

Table of Contents

List of Tables

List of Figures

Acknowledgement

I would like to express my appreciation to my supervisor Andy Johnson for his incredible expertise, time and support – I am truly grateful. Also, to my supervisor Jane Elsley for her time, encouragement, and feedback during the course of this research. I would also like to thank Mike Page and Andrew Paice for their invaluable perspectives and support. I give thanks to the staff at Bournemouth University who gave guidance and to my participants for their time. Lastly, I would like to thank my mum and sister for their unwavering support and belief in me.

1. Introduction

1.1. The Hebb Repetition Effect

The Hebb Repetition Effect (HRE) refers to the finding that surreptitious repetition of a sequence leads to long-term serial learning of that list (Hebb, 1961). Originally, Hebb explained learning and memory from a biological-neuronal perspective in which learning occurs due to changes in neurons within the brain. Specifically, if one neuron repeatedly excites a second neuron, the connection between the two neurons is strengthened and thus fewer errors might be seen in recalling information following repetition (Hebb, 2005). Donald Hebb first tested the effect under a cognitive-psychological viewpoint in an Experiment using Immediate Serial Recall (ISR) of digits. Participants were presented with lists of nine numbers and were tasked with remembering and recalling the lists in the order of original presentation (ISR). Participants were unaware that one list was repeated every third trial (referred to as a Hebb list). When comparing recall performance across the experiment, Hebb found that recall accuracy for the Hebb lists gradually improved over the duration of the experiment. Conversely, recall of the non-repeated (referred to as filler) lists did not increase at the same rate, suggesting the improved recall of Hebb lists were not simply a result of practice effects (see Cunningham et al., 1984 and Fastame et al., 2005 for early replications). Hebb concluded that repetition leads to long-term learning as lists move from the fragile short-term memory to a more durable long-term memory store.

The HRE is particularly important when considering how individuals retain information. The Working Memory Model (Baddeley & Hitch, 1974) described three components within Short-Term memory: the Central Executive, the Visuospatial Sketchpad and the Phonological Loop. The Phonological Loop is where verbal information is rehearsed and temporarily maintained before moving into Long-Term memory and contains two sub-components. The Articulatory Control Process works to refresh verbal information, preventing decay whilst also converting written material into phonological code (Baddeley et al., 1984). This phonological code is then temporarily held in the passive Phonological Store. Under this model, following an adequate number of repetitions, a verbal list would move from the fragile Phonological Loop component to a more durable Long-Term memory store. This model not only accounts for language acquisition (see Baddeley et al., 1998), but also the HRE, suggesting the HRE and language acquisition are functionally similar.

There exists ongoing debate regarding the mechanism underlying the HRE. For example, Burgess and Hitch (1999) argue that serial recall is reliant upon the formation of item-position associations, and that these associations become strengthened following repetition of the Hebb sequence. Such strengthened associations result in each position functioning as an increasingly stronger memory cue for the paired item. However, Cumming et al. (2003) tested Burgess and Hitch's (1999) account by presenting a typical HRE procedure, but with the inclusion of alternating positional 'transfer' lists after eight repetitions of the Hebb sequence. These transfer lists comprised all items from the preceding Hebb sequence, however, only alternate items were in the same serial position as the original Hebb sequence (the remaining items were positioned randomly). For example, if the sequence AHKBJG was presented as the Hebb list, transfer lists could be AGKHJB (or KHABJG), where only every other item is an exact positional replication of the original Hebb sequence. According to the Burgess and Hitch (1999) account, one might predict that the transfer list items presented in the same serial position as the Hebb list should still exhibit a serial recall benefit due to their strengthened position-item associations. This should be observed despite the original Hebb sequence not being represented in full. However, in direct contradiction to this prediction, and, after controlling for protrusion errors by including additional filler lists between the final Hebb and transfer lists, Cumming et al. found no statistical differences in the recall of same-position and different-position items in the transfer list. This suggests that the HRE might not be explained through strengthening of item-position associations. In response, Burgess and Hitch (2006) updated the model setting a threshold for associations to remain active at 60%, tested using simulations. That is, for the strengthened associations to function in the list, at least 60% of the repeated sequence must be reinstated in the transfer list for it to be matched against the previously learnt Hebb list. In the Cumming et al. transfer list, only 50% of the original Hebb sequence was reinstated thus explaining the absence of any recall advantage. Burgess and Hitch (2006) described simulations with the new threshold producing the previously predicted effect that transfer items in the correct serial position were recalled more accurately than items in different serial positions, from the learnt Hebb list. Another update in their 2006 model related to the inclusion of multiple context sets being active at one time (i.e., multiple different list representations can be active). This means that when a new list is presented it can be matched to a number of previous context sets. If a set matches exactly in terms of both item and order information (i.e., an exact repetition) recall is thought to be more accurate due to following the previous context set at recall rather than being represented as a new, less established, context set (Burgess & Hitch, 2006). However, there still exist critiques of the updated model. Burgess and Hitch argue that the strengthening of item and order information, through context sets, occurs after any articulatory processes in the phonological loop (i.e., sub-vocal rehearsal of information to maintain it in the phonological store, and to convert non-verbal stimuli, Baddeley et al., 1984; Baddeley, 1986; Hitch et al., 2009) and subsequently predict that concurrent articulation (CA) manipulations (i.e., repeating words or counting numbers to block the articulatory loop mechanism, Baddeley et al., 1984) would only impact short-term serial recall and not the HRE itself. However, Sjöblom and Hughes (2020) presented evidence of attenuated HREs for auditorily and visually presented sequences when using CA, contrary to Burgess and Hitch's predictions. In summary, Burgess and Hitch (1999, 2006) argue the HRE occurs due to strengthening of item-position associations through matching lists to previous context sets, however, recent evidence from Sjöblom and Hughes (2020), contradicts the updated model's predictions, thus weakening said model as an explanation of the HRE.

An alternative account of what drives the HRE comes from Page and Norris (1998, 2009) who presented a connectionist model, known as the Primacy Model, proposing information such as letters and words are represented as localist units. When a word is presented, for example 'dog', the word unit for dog is activated, as well as the individual letter units ('d', 'o' and 'g'), albeit to a lesser extent as word units are represented at a higher level. Within each unit a primacy gradient exists, comprising of two main components. Firstly, the strength of activation is highest for the first item within the unit (e.g., 'd' in the word 'dog'), with weaker activation strengths in the second, third, fourth items, and so on. This means that when asked to recall the letters of the word dog, 'd' would be recalled first because it has the strongest level of activation. The second component is a suppression mechanism, meaning once an item is recalled it is suppressed so that it cannot be erroneously recalled again. The letter 'd' is therefore suppressed after recall and the next letter with the strongest activation level is recalled (i.e., the letter 'o'). The activation and suppression components work together to ensure that accurate ordered recall occurs. The Primacy Model (Page & Norris, 2009) therefore explains the HRE because when a new list of words such as 'dog, car, red' is presented, individual word units are activated. The primacy gradient mechanism within each word unit relates to letter order so serves no benefit for word recall. Following numerous repetitions of the list and higher-level list unit is formed that combines the individual words. The primacy gradient mechanism within a list unit now relates to the order of the individual words, with the first word, 'dog', having the strongest activation level, which would be suppressed after recall. Subsequent presentations of the list ‘dog, car, red’ would activate the list unit, rather than the individual word units. This explains the gradual improvements in recall seen for repeated lists (i.e., the HRE), as latter presentations have the benefit of the primacy gradient mechanism once the word-items have been chunked together. Non-repeated word-lists are never chunked together to form a higher-level list unit, therefore, never benefit from the primacy gradient promoting ordered recall.

Verbal information such as letters, phonemes and words might not be the only information that can be unitised. For example, units relating to the stress pattern within words could exist (Dell & O’Seaghdha, 1991). This would be of particular importance for homographs (i.e., when two words are spelt the same but have different meanings when pronounced differently) such as, the word Progress. When stress is placed on ‘pro’ the word is a noun, however, when stress is placed on the ‘ess’ the word is a verb. Information regarding the pronunciation or stress pattern of words would likely need to be acquired in order to differentiate between these types of words and could possibly be unitised (M. Page, personal communication, October 3, 2022). Therefore, a future revision of the Primacy model would need to incorporate stress in order to distinguish between verbal stimuli such as the aforementioned homographs.

In summary, the Primacy Model (Page & Norris, 2009) explains the presence of the HRE through chunking and the primacy gradient mechanism. This model also explains how we acquire new words, for example, following numerous presentations of a novel word, letter units are chunked to form a word unit to which we can later assign meaning (Page et al., 2013). Unlike Burgess and Hitch’s (1999) original model, the Primacy Model accounts for data from Cumming et al. (2003) who found that recall for Hebb items in the correct serial position (i.e., matching the previously learnt Hebb list) were no more accurate than Hebb items in a different serial position to the originally learnt list. This finding can, however, be explained by the Primacy Model because it assumes all items are represented as localist units. For example, the Hebb list previously learnt was originally seen as individual letter units that were later chunked together to form a unique list unit. The transfer lists did not match the order of the Hebb list exactly so presentation of these transfer lists would not be considered matches for the Hebb list. They would be considered distinct new lists even though some of the letters were in the same serial position as letter from the Hebb list.

1.2. The Hebb Repetition Effect as an Analogue for Word Learning

The proposition that the HRE is underpinned by the formation of a single list unit in memory (Cumming et al., 2003; Page & Norris, 2009) has similarities with word learning (i.e., the formation of word chunks following lexicalisation). Page et al. (2013) noted three similarities between the HRE and word learning. Firstly, when learning new words, learning occurs despite large intervals occurring between repetitions of the novel words, and consistent with this, Page et al. found that individuals are able to learn new words even when repetitions are infrequent. Page et al. (2013, Experiments 1 and 2) found no significant difference in Hebb learning when the Hebb lists were repeated every third or sixth trial, suggesting, like words, sequences can be learnt with infrequent repetitions. Secondly, when learning new vocabulary, individuals are able to learn multiple words concurrently. To test whether this was also possible with the HRE, Page et al. (2013, Experiment 3) presented participants with a classic Hebb task that included multiple Hebb lists (presented with varying repetition spacings) and found a significant HRE that was not statistically different to Hebb lists learnt in isolation (see also Saint-Aubin & Guérard, 2018). Lastly, Page et al. (2013, Experiment 4) demonstrated that like words, once Hebb lists are learnt, they remain stable over a long period of time. Participants from Experiment 3 were re-tested between three and four months later to investigate the long-term stability of Hebb lists. They found that Hebb lists were recalled with higher accuracy compared to filler lists. However, the effect was shown not to be due simply to familiarity with the sequence items as there was no significant difference in recall between Hebb and filler lists when the order of list items was changed, suggesting that order specific learning of Hebb lists had occurred and remained up to four months after learning. Page et al. therefore concluded that the HRE is a laboratory analogue of word learning.

More recent evidence has also pointed towards a link between language and the HRE. For example, Yanaoka et al. (2019) found that children who scored better on a digit span task (suggesting high phonological short-term memory capacity) showed greater Hebb learning compared to children with low phonological short-term memory capacity (i.e., poorer digit span task performance). This suggests there is a relationship between phonological based memory and Hebb repetition learning, supporting the link between language and the HRE. Moreover, Attout et al. (2020) argued that the mechanisms involved in Hebb repetition learning are important for language-based skills such as reading, based on neural evidence of similar patterns of activation in the hippocampus for people with better reading ability and greater rates of Hebb repetition learning, further supporting the link between language and the HRE.

An additional link between language acquisition and the HRE comes from evidence that Hebb lists can undergo lexicalisation. Szmalec et al. (2009) replicated the classic HRE using sequences of nine letter pairs (split into three sets of three ‘non-word’ pairs). Shortly after the main study, participants completed a Lexical Decision task wherein sets of Hebb and filler pairs were re-presented. Participants showed significantly slower lexical decision times for Hebb compared to filler ‘non-words’, a finding indicative of uncertainty as to whether the non-word was a real word. Szmalec et al. (2009) concluded repetition of the sequences in the main study created long-term phonological representations in lexical memory. This finding suggested that the HRE procedure can induce lexicalisation. To illustrate the durability of this

HRE lexicalisation effect, Szmalec et al. (2012) replicated slower lexical decision making following a 24-hour delay. The HRE-based lexicalisation and long-term stability of the effect provides further support for the link between language acquisition and the HRE.

The final line of support for the link between language acquisition and the HRE comes from studies testing variations in literacy abilities. Firstly, people with developmental dyslexia show both impaired novel word learning (Di Betta & Romani, 2006) and impaired HRE, wherein the proportion of correctly recalled items of repeated lists did not significantly differ from non-repeated lists (Szmalec et al., 2011). This means that dyslexic individuals do not show the same benefit to recall following repetition that is seen with individuals without specific language difficulties. Moreover, Bogaerts et al. (2015) found dyslexic participants (diagnosed with developmental dyslexia during childhood) required, on average, four additional repetitions of a Hebb list to reach the same magnitude of learning, relative to controls. Once lists were learnt, however, retention of lists one month later did not significantly differ between dyslexic participants and controls, suggesting a diminished HRE, as opposed to non-existent Hebb learning. Furthermore, Bogaerts et al. (2016) found a positive correlation between serial-order learning and reading, in that children with poorer reading ability showed weaker rates of Hebb learning. They suggested that holding temporary representation of order (and the ability to consolidate said sequential information in a long-term store) are both implicated in language acquisition and reading. Evidence from both Szmalec et al. (2011) and Bogaerts et al. (2015; 2016) point towards a diminished HRE in people with language and reading difficulties, supporting the view that the HRE could be a laboratory analogue of word learning.

There is, however, also evidence suggesting that people with language difficulties have difficulties with serial memory in general, rather than a diminished HRE. This means that rather than there being a reduction in the gradual recall improvements seen for Hebb list only, there is simply a reduction in recall accuracy across all trials. For example, illiterate individuals show significantly poorer serial recall, in general, relative to literate individuals but still show a HRE (Smalle et al., 2019). Similar findings have also been shown in adolescents (aged 11-15 years old) with non-specific intellectual difficulties showing a HRE that was consistent with the accuracy shown by typically developed 4–10-year-olds (Henry et al., 2022). Mosse and Jarrold (2010) also found children with Down's Syndrome showed increased recall errors across all trials relative to typically developed controls alongside evidence of Hebb repetition learning. Whilst these studies point towards a view that the HRE is not diminished in the above populations and rather there is a general deficit in serial memory, Henry et al. (2022) and Mosse and Jarrold's (2010) samples show general intellectual difficulties, not just language difficulties. When looking at studies with specific language difficulties (i.e., Dyslexia and reading ability, Bogaerts et al., 2015; 2016; Szmalec et al., 2011), there is evidence of a diminished HRE. In fact, Bogaerts et al. (2016) noted that future nonword reading ability was significantly predicted by Hebb learning, whilst filler list performance did not significantly predict nonword reading. This suggests there is a relationship specifically between the HRE and reading ability and not a broader relationship between serial memory and reading ability. Taken together, these studies highlight a strong link between the HRE and word learning and that the HRE can be seen as a laboratory analogue of how we learn new words. The HRE is therefore an appropriate paradigm to investigate how we learn sequences of verbal information, for example, in the present research, whether repetition of semantic patterns alone (as opposed to exact item repetition) can produce long-term learning of lists.

1.3. Cross-Modal Hebb Repetition Learning

While a growing body of evidence supports the HRE as a laboratory analogue of word learning, it is important to note that the HRE has also been found with a range of non-verbal stimuli suggesting the effect is not uniquely verbal. The Working Memory Model (e.g., Baddeley & Hitch, 1974; Baddeley, 2000) provides a framework through which cross-modal memory can be viewed (although it is important to note that this is by no means the only framework of short-term memory). In this model, memory for verbal, visuo-spatial and other stimuli are stored in separate slave systems (the Phonological Loop, Visuo-spatial Sketchpad and Episodic Buffer, respectively) under the attentional control of the Central Executive. If the HRE is purely a verbal effect associated with language acquisition, under this model, the effect would be a purely Phonological Loop based phenomenon and no cross-modal effects would exist. Page et al. (2006) investigated whether the HRE relied solely on the phonological loop by including a condition of CA in a Hebb repetition task. CA is a secondary verbal task that participants complete during learning and/or recall, in order to block the use of the phonological loop for the main task, such as counting backwards or repeating a specific word. It is thought that CA prevents the conversion of graphemic stimuli to phonological code (Baddeley et al., 1984). For example, if presented visually with the word ‘house’, CA (e.g., repetition of the word ‘the’) would disrupt the ability to covert the word ‘house’ into phonological code and prevent sub-vocal rehearsal. In their first experiment, Page et al., (2006) administered a typical HRE task using lists of English letters, with the inclusion of participants repeating the word ‘racket’ throughout presentation and recall of the lists. They found a significant HRE regardless of the addition of CA, suggesting the HRE persists without access to the phonological loop. In a second experiment Page et al. used sequences of object images (such as bikes or houses) and again found significant HREs, with learning surviving an additional CA condition. Notably in their second experiment, the effect disappeared with visual stimuli, under CA, if Hebb lists included category repetitions (i.e., each repetition of dog or bike would be a different breed of dog or style of bike, respectively), as opposed to identical repetition (i.e., the exact same images being repeated). There was however, a HRE for category repetitions under the tapping secondary task (a secondary task thought to block the use of the Visuospatial Sketchpad). Page and colleagues suggested this occurred because participants were likely naming each image of a dog, as dog, rather than the specific breed, most likely because there was no specific task instruction to label images in any greater detail. In the CA condition, participants were prevented from sub-vocal rehearsal of ‘dog’, therefore, no exact repetition occurred throughout the experiment, resulting in no Hebb learning. However, in the tapping condition, the phonological loop was unoccupied enabling participants to sub-vocally rehearse lists, most likely rehearsing the broader category names, resulting in significant Hebb learning. This experiment and its implications will be discussed further in relation to the present research later (see Section 1.6 The Present Research: Semantic Information and the HRE). Overall, and according to a Working Memory Model conceptualisation, the above evidence suggests the HRE is not purely a verbal effect, and therefore not uniquely reliant upon the phonological loop.

The HRE has been found with a range of other non-verbal stimuli. For instance, Couture and Tremblay (2006, see also Guérard et al., 2011) used a similar experimental procedure as Hebb (1961), however, they presented participants with sequences of seven dots in different spatial locations, with the Hebb list being repeated every fourth trial. At test, participants were required to click on the locations of the dots in the order of original presentation (i.e., a procedure called Serial Order Reconstruction: SOR). Similar to verbal effects, the rate of improvement was significantly greater for the Hebb compared to filler sequences. Moreover, in a second experiment, where Couture and Tremblay directly compared verbal and visuo-spatial sequence learning, they found no significant difference in mean improvement rates for repeated sequences across stimulus types. This learning equivalency tentatively suggests that the HRE is functionally similar for verbal and non-verbal stimuli, rather than a modular feature of verbal memory. However, cross-modal differences were found by Sukegawa et al. (2019, Experiment 2), wherein random temporal grouping (i.e., inserting non-repeated pauses in between serial positions of the Hebb list) did not weaken the HRE when using sequences of dots in different spatial locations. This is inconsistent with evidence using verbal stimuli in which random temporal grouping weakens (Hitch et al., 2009) or abolishes (Bower & Winzenz, 1969) the HRE. Sukegawa et al. (2019) concluded some evidence does point towards domain-specific, as opposed to domain-general mechanisms in the HRE.

The cross-modal generality of the HRE is, however, further supported by Horton et al. (2008) who found a significant HRE using sequences of unfamiliar faces (see also Johnson et al., 2017; Johnson & Miles, 2019a,b). Horton et al. did not, however, find a HRE when using inverted faces, which they concluded to reflect low levels of psychological distinctiveness when facial stimuli is inverted (i.e., inverted faces are deemed less distinct due to lesser experiences relative to upright faces that are more distinct, Hay et al., 2007). Johnson and Miles (2019a) described several reasons why Horton et al. (2008) found no HRE in their study, including there not being enough repetitions of the Hebb list, scoring criteria that was too strict, and an insufficient number of participants to detect the HRE. When these limitations were rectified, Johnson and Miles (2019a) found a HRE for both inverted faces and matrices, suggesting the HRE can be found across multiple modalities (even with stimuli possessing low levels of psychological distinctiveness). To further illustrate the cross-modality of the effect, the HRE has been shown with the spatial position of visual (Mosse & Jarrold, 2008) and auditory information (Parmentier et al., 2008), tactile stimuli (Johnson et al., 2016), and odours (Johnson et al., 2014). These findings show that whilst the HRE might help understand language acquisition, the HRE is a general feature of memory and not confined to verbal memory.

1.4. The Role of Stimulus Overlap and Repetition Spacing on the HRE

Recent research has focused on investigating factors that influence the HRE. For example, item overlap, which refers to the items used to construct Hebb and filler lists, has been explored as a factor that can affect the rate of learning for the repeated sequence. Evidence suggests that when Hebb and filler lists are comprised of the same items, the HRE is significantly reduced, compared to lists comprised of different items (full overlap vs. no overlap), in verbal (Page et al., 2013; Smalle et al., 2016) and visual modalities (Johnson et al., 2017; Johnson & Miles, 2019a). This has some further analogue with language acquisition where letters used to create words are sampled from a large stimulus set (for example, 26 letters in the English alphabet enabling reduced stimulus overlap across words). However, Hebb's (1961) original study used Hebb and filler lists comprised of the same nine numbers (full overlap) yet still found the classic HRE. Furthermore, St-Louis et al. (2019) found, with a large sample (N=210), that fully overlapping items could still produce a significant HRE, demonstrating that the effect is not completely abolished when repeated and non-repeated items overlap. The role of repetition spacing has also been investigated, which refers to the number of filler lists presented between Hebb lists. Hebb's (1961) study used two intervening filler lists, with Melton (1963) suggesting that the HRE was reduced with 3-5 intervening lists and that no HRE was detected with longer spacings. However, more recent evidence has shown significant Hebb learning with 5-11 intervening filler lists in both verbal and non-verbal modalities (Johnson et al., 2017; Page et al., 2013; Page & Norris, 2009; St-Louis et al., 2019). This suggests that factors such as stimulus overlap may interact with repetition spacing in determining both the presence and magnitude of the HRE.

1.5. Additional Pattern Structures that affect the HRE

The previous evidence discussed above provides strong evidence that item and order information are acquired following repetition. However, researchers have recently investigated if other information is also acquired during sequence learning. For example, Johnson et al. (2017) investigated whether repetition of motor responses during recall had an additive benefit on Hebb learning. They administered a Hebb repetition task in which participants memorised lists of five unfamiliar faces. At test the five faces were shown simultaneously on the screen and participants had to click each face in the order of original presentation to reconstruct the list (SOR). Motor responses were manipulated by changing the location of faces on the test screen. For Hebb lists, half of the participants saw a recall screen that had the items in the same spatial location on the screen. This meant that exact motor responses (from clicking each item) were also repeated along with the item order typically seen in a Hebb trial. The remaining participants saw recall screens in which the spatial location of faces were always novel. This manipulation meant that the item and order information were repeated, however, the motor responses required to reconstruct the list were not. Johnson et al. (2017) found that when motor responses were repeated across the experiment, the HRE was larger compared to novel motor responses (see Fendrich et al., 1991 for additional evidence of motor response HREs). Although a HRE was still found when motor responses were novel, repeated motor responses were found to have an additive benefit to long-term serial order learning. This suggests that while the effect is not merely underpinned by motor response outputs, there is some additional learning of motor response patterns following repetition.

Information about the metrical pattern of verbal items might also be acquired following repetition. A metrical pattern refers to the rhythm in which a word is articulated, specifically relating to the number of syllables in a given word and the placement of the stress (sometimes referred to as ‘accent’) within a given word. For example, the word ‘popular’ is a three-syllable word with the stress on the first syllable, whereas ‘consider’ is a three-syllable word with the stress on the second syllable. Paice et al. (in preparation) administered a SOR task where lists of words were presented auditorily. Unlike the classic Hebb paradigm where the items within a Hebb-list are kept constant over subsequent repetitions, in this novel experiment the items differed across Hebb-list repetitions but crucially the underlying metrical pattern was instead kept constant. For example, if one Hebb-list began with the words: “donkey…house… addition”, the second Hebb-list might begin with the words: “monkey…king…condition”. Within a given block, they presented nine such lists comprising the same metrical pattern, followed by a single filler list comprising of novel items and a novel metrical pattern. They found a significant learning gradient, above what would be expected from practice, suggesting it is possible to learn metrical patterns through repetition. As previously mentioned, under the Primacy Model (Page & Norris, 2009) verbal HREs are explained by the presence of phoneme units. There also exists the possibility that stress/strength units might exist, to explain the distinction between homographs (M. Page, personal communication, October 3, 2022). Whilst not currently addressed in any iterations of the Primacy Model, it’s possible that the architecture could be extended to include strength units that hold information regarding the metrical pattern of words (e.g., strong, weak, weak for the three-syllable word ‘Popular’, M. Page, personal communication, October 3, 2022). Under this view, the HRE in the Paice et al. study, might have been found without the need for exact item repetition because repetition of the same metrical patterns resulted in the formation of a singular chunked metric unit, with activation of this unit facilitating list recall. However, this remains rather speculative and would certainly require additional evidence and explanation in future revisions of the Primacy Model.

The above evidence points towards a view that additional information, other than precise item and order information, is obtained during learning and retrieval of repeated lists. Specifically, repeated motor responses (Johnson et al., 2017) and metrical patterns (Paice et al., in preparation) have been shown to produce an additive benefit to Hebb learning. If these, and potentially other factors, do indeed produce a stronger Hebb effect, this could be used to help maximise long-term sequence learning.

1.6. The Present Research: Semantic Information and the HRE

The focus of the present research surrounds whether semantic information is acquired during long-term sequence learning. Here, analogous to Paice et al. (in preparation), we test whether repeating the same pattern of semantic information results in improved serial recall, despite the words used to construct those sequences being experimentally novel. The main line of evidence that points towards the possibility of semantic information being acquired alongside item and order information comes from the Semantic Similarity effect. Multiple studies (Poirier & Saint-Aubin, 1995; Saint-Aubin & Poirier 1999; Saint-Aubin et al., 2005) have shown that when you present lists in an ISR task, lists comprised of semantically similar items (e.g., Cat, Dog, Horse, Cow) are recalled significantly more accurately than lists comprised of semantically dissimilar items (e.g., Cat, Blue, House, Pear). Saint-Aubin et al. (2005) explains the effect in terms of a shared semantic link between items providing an additional retrieval cue, and subsequent recall advantage, compared to lists without a semantic link. Botvinick and Plaut’s (2006) model of serial recall posits that background knowledge in general impacts serial recall accuracy. They put forward a connectionist model of serial recall in which connection weights are adjusted gradually after detection of regularities in list structures to improve serial recall. Whilst this model does not attempt to account for long-term serial order learning, it does suggest that consistencies in sequential structure impact encoding for serial order memory tasks. If required, it could be possible to extend this model to explain the role of background knowledge on long-term serial order memory.

Based on previous evidence of a semantic similarity effect in serial order memory, one might posit that this reliable pattern of semantic information may provide additional cues for the to-be-remembered sequence, reducing errors and facilitating the rate at which the sequence is acquired (i.e., extending to long-term serial order memory). However, whilst Page and Norris' (2009) model accounts for long-term serial order learning, it would not predict a HRE for semantic information. This is because the formation of the chunks in the model (i.e., conversion of item strings into an individual unit) is reliant on exact replications of the items in the same order. This follows from the same 'items' being repeatedly recalled and then suppressed at the same point in the list resulting in a more distinct activation level within the primacy gradient of the sequence. Here, whilst the semantic pattern is repeated, the localist item information differs for each repetition of the Hebb sequence. An exception to this rule, as stated previously, is that it is possible to show metrical Hebb repetition learning (i.e., repetition of metrical patterns without exact item repetition, Paice et al., in preparation); explained by the possibility that the strength of syllables are detected and unitised in a similar way to phonemes (M. Page, personal communication, October 3, 2022). Although it seems somewhat plausible that information surrounding the strength or stresses within words could be unitised, it differs from semantic information, in that our understanding or detection of semantics might require a higher order cognitive process. Therefore, it might not be as easily explained by a model, such as the Primacy Model, which deals with lower-level features of words (M. Page, personal communication, October 3, 2022).

The present research chose to focus on whether there is a Hebb effect for semantic information (i.e., repetition of semantic categories rather than exact items) rather than how semantically similar or dissimilar lists influence the classic HRE, akin to the semantic similarity effect literature. The former would utilise semantically heterogenous lists such as Apple, Blue, Dog with the latter utilising homogenous lists such as Apple, Pear, Grape. This decision was made because investigating if there is a Hebb effect for semantic information, follows on more closely to the work of Paice et al. (in preparation) who investigated whether HREs can be seen without exact item repetition, assuming another aspect of the list is repeated. Furthermore, the use of heterogenous lists, as opposed to homogenous, is more closely related to Page et al. (2006, Experiment 2) and a more appropriate choice of stimuli to investigate whether there is a HRE for semantic information. If the same effect exists with semantic information (as shown by Paice et al., in preparation) this would prompt revisions of prominent HRE models (Burgess & Hitch, 2006; Page & Norris, 2009) and is therefore important to investigate.

As previously discussed, Page et al's. (2006) Experiment 2 links closely to the present research. Most importantly, they found no Hebb learning for category repetitions (repeated category names but different images in each Hebb list) under conditions of CA (repeating the word racket during encoding of the list). This points towards there potentially not being a semantic Hebb effect, i.e., Hebb learning with repetition of semantic/category patterns, and not exact item repetition. If semantic information was able to be unitised, there would have likely been a Hebb effect for category repetition even under CA, in line with typical HREs. However, the main aim of their research was not specifically to test whether a semantic Hebb effect exists, therefore, their experiment might not have been designed in a way to maximise finding a semantic Hebb effect, if it exists. For example, the ‘categories’ used were low-level (e.g., houses, dogs, bikes and so on), as discussed previously, participants were less likely to view different breeds of dogs as different items and would have likely used a strategy to sub-vocally rehearse dog for every list – a strategy that was blocked by using visual stimuli alongside CA. Subsequently, the ability to measure a semantic Hebb effect might have also been blocked, because participants would not have ever viewed the category lists as distinctly different lists. Using verbal stimuli and well as higher-level (and thus more distinct) categories might facilitate finding a semantic Hebb effect, if it does indeed exist.

The present research aims to investigate whether semantic information is acquired during repetition learning, using a Hebb repetition paradigm. This will help identify whether semantic information can act as an additive factor in the Hebb repetition effect, and subsequently inform ways to maximise long-term serial order learning. Furthermore, this research will test the appropriateness of current models of the HRE which are confined to lower-level features of words influencing verbal Hebb learning, as opposed to higher level features such as semantic information. Based on evidence that additional information can be acquired alongside item and order information (Johnson et al., 2017; Paice et al., in preparation), and evidence of the semantic similarity effect (Poirier & Saint-Aubin, 1995; Saint-Aubin & Poirier 1999; Saint-Aubin et al., 2005), this study predicts that semantic information can be acquired during repeated serial order learning, which will be tested across three experiments.

2. Experiment 1

The first experiment was designed to test whether it was possible to achieve a Hebb repetition effect when the Hebb lists are not comprised of the exact same items, but the same semantic pattern is repeated. For example, if the Hebb pattern follows a semantic sequence of ‘Animal, Sport, Fruit, Colour’, the first Hebb list might be ‘Dog, Golf, Apple, Blue’ and the second would be ‘Cat, Tennis, Pear, Red’. Following a similar methodology to Paice et al. (in preparation), Experiment 1 will present a series of lists that comprise of novel words that share the same semantic pattern. In each block participants will receive 9 semantic Hebb lists, 2 filler lists, and a transfer list. The semantic Hebb sequences are all unique but share the same order of categories. As in Paice et al., these semantic Hebb lists are presented concurrently in order to maximise the opportunity to detect learning. The filler lists are all unique and have a different order of semantic categories. The transfer lists operate as an additional test of item-position strengthening (as undertaken by Cumming et al., 2001 to test the predictions of the Burgess & Hitch, 1999 model). Here only alternate items in the transfer list repeat the semantic categories in the same position as in the semantic Hebb list. If participants learn the order of the repeated semantic pattern, it is predicted that there will be a significant reduction in recall errors for lists presented later in the block, i.e., a semantic HRE. Following such an effect, one might also predict significantly lower error rates for the final Hebb list compared to the filler lists (novel list items following a novel semantic pattern). To additionally test whether any semantic HRE is merely a carryover effect from the preceding semantic Hebb trial, we manipulate whether or not the final Hebb sequence is preceded by an intervening filler sequence. This means some participants will receive all nine Hebb lists, then filler lists (‘filler after the final Hebb’), or eight Hebb lists, one filler and the final Hebb list (‘filler before’ the final Hebb). Finally, if any semantic HRE is underpinned via the strengthening of item-position associations, one might predict significantly lower error rates for the transfer list compared to the filler lists. Experiment 1 therefore has three hypotheses. Firstly, there will be a gradual reduction in errors made when reconstructing semantic Hebb lists, indicating a Hebb effect for semantic information. Secondly, there will be no significant difference in errors made on the final Hebb list when it is presented immediately after the eighth presentation, or after an intervening filler list, suggesting the effect is not attributed to carryover effects. Thirdly, there will be no significant difference in errors between transfer list performance (i.e., lists in which only alternate items in the list are repeated from the Hebb list, analogous to Cumming et al., 2003) and filler list performance, supporting a primacy gradient explanation of the effect (Page & Norris, 2009), as opposed to an item-position association explanation (Burgess & Hitch, 2006).

2.1. Method

2.1.1. Design

A mixed multifactorial (3x2) design was employed where the within-participants factor was trial type (Hebb, filler, and transfer) and the between-participants factor was position of the first filler sequence (before or after the final Hebb list in the block). The dependent variable is serial reconstruction accuracy. As stated in the pre-registration (https://osf.io/wsyv8) an additional analysis was conducted with a repeated measures independent variable of trial type (Hebb and filler) and a dependent variable of learning gradient.

There were six experimental blocks. Block 1-5 contained 9 Hebb sequences, 2 filler sequences, and 1 transfer sequence. Blocks 1-5 each had a unique Hebb sequence. In order to assess longevity of any learning, half the participants in Blocks 1-5 received the first filler sequence before the final Hebb sequence and half of the participants did not. To further explore the durability of any Hebb repetition learning acquired in Blocks 1-5, Block 6 re-tested some of the sequences from the preceding blocks. Block 6 contained five Hebb and five filler sequences (one Hebb and one filler sequence were sampled from each of the preceding blocks). These Hebb sequences were comprised of experimentally unique items but followed the semantic Hebb patterns used in Blocks 1-5.

The Hebb sequence comprised a repeating pattern of semantic categories. Specifically, whilst the items used to construct each Hebb sequence within a block were different, the same order of semantic categories within each Hebb list was repeated. The filler sequences comprised a unique order of the semantic categories, with both filler sequences in a block comprised of different words. The transfer list comprised 3 category words in the same position as the Hebb sequence and 3 category words in a novel order. The repeating elements of the Hebb sequence were positioned in alternate serial positions in the transfer list and the order of the alternating repeating/non-repeating items was counterbalanced (i.e., the Hebb items were either at positions 1, 3 and 5, or positions 2, 4, and 6).

2.1.2. Participants

A power analysis was run using a Generalised Linear Mixed Effects Model with a probit link function, using R (https://www.r-project.org/). The model had a single fixed effect, which was an unstandardised effect size of trial typeset at 0.18. The analysis revealed that to obtain 80% power with an alpha level of .05 and an effect size of 0.18 we would need to collect 120 participants.

Participants were recruited via two platforms. Psychology students (*n* = 143, *M* = 20.57, *SD* = 3.95, aged 18-46, 120 females, 2 'Prefer Not to Say') from Bournemouth University were recruited through Sona Systems (https://www.sona-systems.com) and participated in exchange for course credits. Additionally, English-speaking, non-students (*n* =113, *M* = 27.04, *SD* = 8.3, aged 18-50, 37 females) were recruited through the online recruitment website ‘Prolific’ (www.prolific.co) and were paid £3.75 as compensation. Participation in the pilot study (see Appendix A1 and A2) was an exclusion criteria and participants were required to be aged between 18-50yrs.

Due to the pre-determined performance criteria (see ‘2.1.5. Data Analysis’), 136 participants (75 students and 61 non-students) were excluded from the analysis. The final sample, therefore, consisted of 120 participants (*M* = 23.57, *SD* = 7.46, aged 18-49, 81 females), including 68 students (*M* = 20.59, *SD* = 4.27, aged 18-46, 58 females) and 52 non-students (*M* = 27.46, *SD* = 8.87, aged 18-49, 23 females). The age of participants was not normally distributed ($W = 0.71$, $p < .001$) with 18 age-related outliers and a skew towards a younger adult age.

2.1.3. Materials

The experiment building software ‘Gorilla’ (build 20201002, https://gorilla.sc/) was used to create and run the present study (Anwyl-Irvine et al., 2020). Seventy-eight single syllable 4-letter words were used as stimuli. Stimuli was generated from a pilot study in which participants (n = 50) selected the best 13 category exemplars (from a set of 20 words) and repeated this process for six categories (see Appendix A1 and A2 for details of Pilot Method and Results). Words were spoken by a male voice, recorded and edited using the application ‘Audacity’ (version 2.4.2, https://www.audacityteam.org/). Words were cut into 500ms sound files. Stimuli were grouped into six categories (Animals, Body Parts, Food, Music, Names and Places) each containing 13 words. Twenty semantic patterns were randomly generated to make the five Hebb and 15 novel filler patterns (5 Hebb patterns and 10 filler patterns were used for blocks 1-5 and the remaining five filler patterns were used in block 6). There were also five transfer semantic patterns that included the same categories as the Hebb list but arranged with three items in the Hebb serial position and three items in novel serial positions (this was randomised across study versions). Four versions of the study were generated so that every semantic pattern was used once as a Hebb pattern. Lists were comprised of one word from each of the six semantic categories, these compositions were also novel for each version of the study.

Overall, there were eight versions of the study to which participants were randomly assigned. Four of the experimental versions are unique, each including different list compositions. The remaining 4 versions were identical except for the positioning of the first filler sequence in Blocks 1-5. In the first four versions, participants are presented with nine Hebb lists followed by the first filler list and in the other four versions, participants are presented with the first filler list before the final Hebb list.

The same set of 78 words were used in every block but each block had a novel set of list orders, with no two words appearing in the same list more than once. For example, if ‘Goat’ and ‘Jack’ appeared together in a list, they would not appear together in the other blocks of that specific experimental version.

In the sixth block, the unused stimulus set words from each block (one word per semantic category) were used to re-present the semantic Hebb sequence from that block. These five Hebb sequences (one from each block) were randomly presented with five novel filler lists.

2.1.4. Procedure

Participants completed the study online. After obtaining informed consent participants were asked for their age and gender as demographic information. They were instructed to use headphones and had an opportunity to test the sound before the main task started. The main task involved participants being auditorily presented with six words, sequentially. Auditory presentation of stimuli was chosen, as opposed to visual presentation, to prevent participants from simply looking at the first letter of each item and sub vocally rehearsing that. Auditory presentation would encourage processing of the whole word more so than visual presentation. Also, auditory presentation was chosen to replicate Paice et al's. (in preparation) method as closely as possible.

	Animal	**Body**	**Food**	**Music**	**Name**	**Place**
Hebb	Goat	Foot	Cake	Jazz	Beth	Rome
Hebb	Bear	Knee	Eggs	Drum	Jack	Kent
Hebb	Wolf	Hand	Rice	Song	Kate	Hull
Filler	Band	York	Jess	Deer	Head	Pork
Transfer	Bull	Hyde	Soup	Hymn	Luke	Neck
Transfer	Stew	Harp	Josh	Nose	Swan	Bath

Figure 1. Experiment 1 list examples.

After a 3000ms interval, the six words were presented sequentially in 500ms sound envelopes (i.e., each word was contained within a 500ms audio file, regardless of the length of the word), with a 500ms inter-stimulus-interval between each word. At test, following a 500ms retention interval, all six words were presented visually to the participant on the screen in a randomised order. Participants were required to click on each word in the order that they heard them. Words could only be selected once and disappeared when selected. There was no time limit for reconstructing the list. Participants completed 70 of these trials (see Figure 1 for list examples), with the commencement of each trial self-paced. On completion of the task, participants were asked three questions regarding the lists they had just learnt. Firstly, participants were asked whether they noticed anything "special" about the lists they had learnt. Secondly, whether they noticed any repetition in the lists. Thirdly, participants were asked to elaborate on what they thought had been repeated. Participants typed their responses to questions one and three, and a button response (yes or no) was used for question two. The study took approximately 30 minutes.

2.1.5. Data Analysis

There were two pre-registered performance-based exclusion criteria (https://osf.io/wsyv8). First, if the overall average serial recall accuracy for the filler lists and first Hebb list of each block (the first Hebb list of each block is a de facto novel sequence) was 85% or above (i.e., less than .15 proportion serial recall errors), the participant was excluded, as near ceiling performance would prevent the detection of any Hebb repetition learning. Second, if the overall average serial recall accuracy for the filler lists and first Hebb list of each block (i.e., the novel unrepeated sequences) was 65% or lower (i.e., greater than .35 proportion serial recall errors), the participant was excluded. This is based on a score of 66% being obtainable if the participant focussed on remembering 3 items in the list and guessing the remaining items.

Learning of the repeated semantic pattern was analysed using generalised linear mixed effect models (GLME) in R Studio (R Studio Team, 2021) using the lme4 package (version 1.1.26). Following Barr et al's. (2013) procedure, the largest random effect structure that converged would be reported. The model used the fixed effects of trial type (final semantic Hebb list in a block and first filler in a block) and first filler position (before or after the final semantic Hebb sequence). A random slopes model was used as a function of stimulus item for the fixed factor of trial type and random intercept effects were used for participants. The dependent variable was serial recall accuracy at each epoch collapsed across the first five blocks.

Analysis was conducted using a Generalised Linear Mixed Effects Model. This analysis is argued to reduce the chances of Type 1 errors based on accounting for both by-participant and by-item variance (Singmann & Kellen, 2019), as well as dealing with non-normal data better than General Linear Models (Bolker et al., 2008). Analysis was also conducted using General Linear Models (i.e., *t*-tests etc.) alongside the Generalised Linear Mixed Effects Model analysis. This allowed for direct comparisons between the results of two types of statistical analysis. If both methods of analysis produce consistent results across experiments, it provides evidence that comparisons could be made across papers where different methods of analysis have been chosen. Moreover, it supports the appropriateness of using Generalised Linear Mixed Effects Model analysis in future HRE experiments, if it produces the same broad findings as General Linear Model analysis, which to-date has been the more typically used method of analysis for HRE literature (such as in experiments from Hebb 1961; Johnson et al., 2013; Johnson et al., 2016; Mosse & Jarrold, 2010; Norris et al., 2018; Page et al., 2013, to name a few). Finally, Bayes factors were also reported alongside the General Linear Model analysis to provide an additional measure of effect strength, using the typical cut offs for substantial evidence of the alternative hypothesis ($BF_{10} = 3$) or null hypothesis ($BF_{01} = 0.33$, Bolstad & Curran, 2016).

2.2. Results

2.2.1. Descriptive Data

The data, as seen in Figure 2, showed the typical serial position curve expected with auditorily presented lists, showing a strong primacy and slight recency effect for both the Hebb and filler trials (Drewnowski & Murdock, 1980).

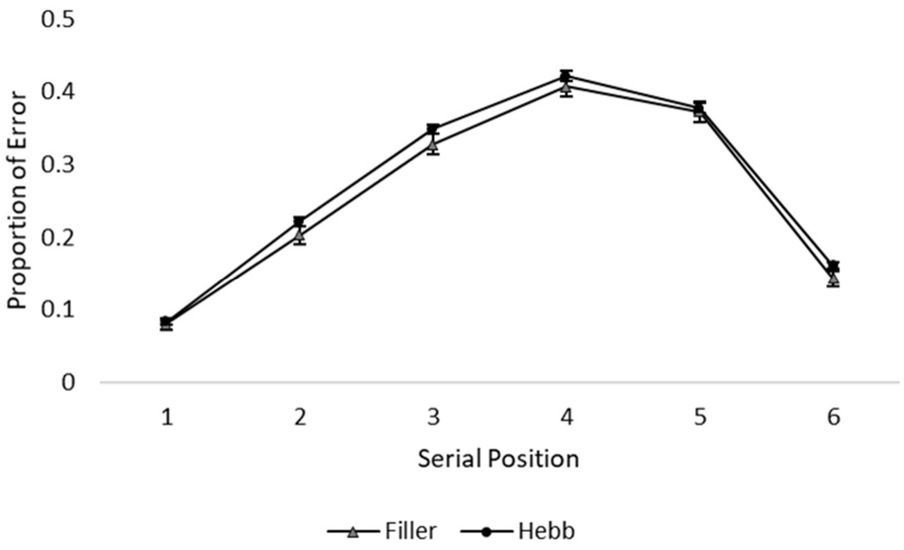


Figure 2. Mean proportion of errors made for each serial position for both the Hebb and filler trials, collapsed across all blocks. Error bars represent standard error.

Data from blocks one to five were collapsed and analysed together (see Figure 3 for mean proportion of errors for blocks one to five). Figure 3 shows an almost straight mean error gradient of -.001, with overlapping error bars, suggesting there was likely no significant learning across the experiment. See also the bar in Figure 3 showing the proportion of errors for the filler lists which show similar performance in comparison to Hebb lists, again, suggesting non-significant Hebb learning.

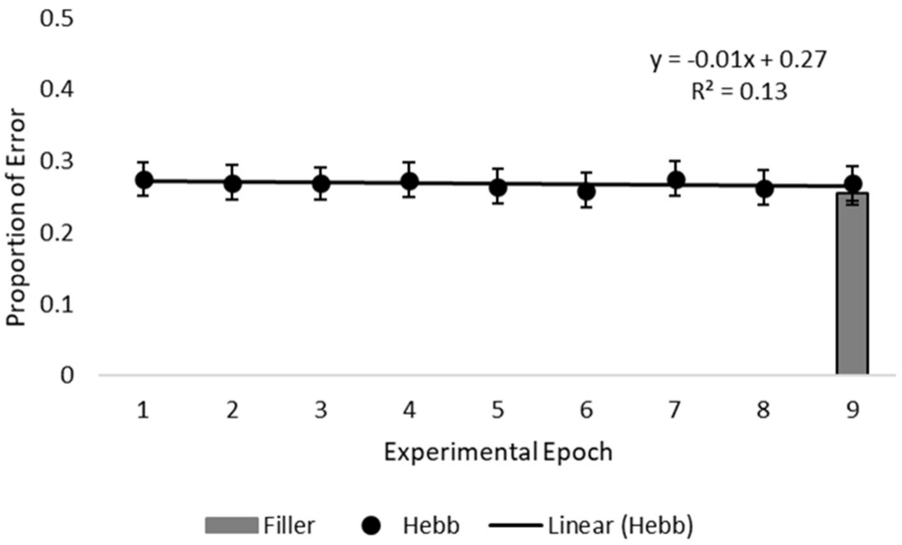


Figure 3. Mean proportion of error for each repetition of the Hebb list as well as the mean error for filler lists, collapsed across block 1-5. Gradient equation and R-squared values are noted for Hebb and filler lists. Error bars represent standard error.

Descriptive analysis was conducted regarding the awareness questions at the end of the study. When asked whether they noticed anything being repeated, 103 of the 120 participants responded ‘yes’. However, when asked to explain what was repeated only two participants noted that there were some lists that followed the same semantic pattern. Most simply stated that words were repeated across blocks or listed specific words that they noticed appearing multiple times across the study.

2.2.2. Generalised Linear Mixed Effects Model

The GLME model is reported in Table 1 and includes estimates, standard errors, z-values, and p values. Importantly there was a non-significant fixed effect of trial type, indicating that the difference between the average serial recall for the final semantic Hebb sequence and the first filler sequence did not significantly differ. There was no overall effect of the filler position on recall and nor did this interact with trial type. The analysis indicates no evidence of improved recall for the semantic repetition. Based on the non-significant findings, no analysis was conducted on block 6. This block was included to investigate the durability of Hebb learning by presenting previous Hebb and filler lists after the initial blocks. Given no learning was found, formal analysis of Hebb (*M* = .22, *SD* = 0.25) and filler (*M* = .25, *SD* = 0.26) errors in block 6 were not necessary. Similarly, no analysis was conducted on transfer trials due to a non-significant Hebb effect.

Table 1.

Coefficients from the GLME analysis

	Estimate	Std. error	z-value	p-value
Intercept	1.12	0.09	12.50	< .01
Trial Type	-0.13	0.08	-1.59	.11
Filler Position	-0.06	0.12	-0.52	.60
Trial Type*Filler Position	0.18	0.11	1.62	.11

2.2.3. General Linear Model

The General Linear Model analysis included two independent variables, Trial Type (Hebb and filler) and Filler Position (before or after the final Hebb list). The dependent variable was mean proportion of errors made. Hebb and filler data were collapsed across Blocks 1-5. The assumptions of normality (all W > .98, all $p > .068$) and equality of variance (all F > .47, all $p > .366$) were met for all conditions. A mixed 2x2 ANOVA (Trial Type x Filler Position) found that there was no significant main effect of Trial Type, *F*(1, 118) = 0.21, *MSE* = 0.01, $p = .646$, $\eta_p^2 < 0.01$, $BF_{10} = 0.15$, or Filler Position, *F*(1, 118) = 0.1, *MSE* =

0.02, $p = .751$, $\eta_p^2 < 0.01$, $\text{BF}_{10} = 0.18$. The interaction between Trial Type and Filler Position was also not significant, $F(1, 118) = 1.36$, $MSE = 0.01$, $p = .246$, $\eta_p^2 = 0.01$, $\text{BF}_{10} = 0.01$.

2.3. Discussion

The aim of Experiment 1 was to establish whether semantic information is acquired alongside item and order information by repeating a semantic pattern (i.e., without exact item repetition) in a Hebb repetition study. Experiment 1 followed the same methodological procedure as Paice et al. (in preparation) who, contrary to the present experiment, found evidence to suggest the HRE can be produced without exact item repetition. In their study they found evidence that metrical information (i.e., repeating sequences of items following the same word-stress patterns) produced a typical HRE.

Results from both the Generalised Linear Mixed Effects model and the General Linear model analyses showed that there was no significant reduction in recall error for the final repetition of the semantic patterns compared to the novel filler patterns. This means that, contrary to the hypothesis, repetition of a semantic pattern alone does not produce the typical Hebb repetition effect, wherein the final repetition of the Hebb list would have significantly fewer errors, relative to filler lists (Hebb, 1961; Johnson et al., 2017; Page et al., 2013). There was also no significant effect of filler position, meaning that recall errors for the final repetition of the semantic Hebb pattern were not affected by the presence of an intervening filler list. This is a seemingly logical finding given the non-significant Hebb learning. As stated in the pre-registration (https://osf.io/wsyv8) and in the Results section, analysis on the final block, which involved presentation of previous semantic Hebb and novel fillers, was not conducted due to non-significant semantic pattern learning.

It is worth noting that the absence of Hebb repetition learning is unlikely to have resulted from online testing, wherein there is less control over participation and accurate completion. There is a threefold justification for this. Firstly, Figure 2 shows typical serial position curves for auditorily presented lists, suggesting participants completed the task properly. Secondly, the large sample size (n = 120) would likely mitigate additional noise in the data resulting from a less controlled procedure, relative to in-person testing. Thirdly, Paice et al. (in preparation) found a significant HRE with metrical patterns using a highly similar online procedure to Experiment 1.

Overall, results from Experiment 1 show that repetition of semantic patterns alone do not produce Hebb learning, thus pointing towards the view that semantic information is not acquired alongside item and order information in the same way as metrical patterns (Paice et al., in preparation) or motor responses (Johnson et al., 2017) might be. This is consistent with the Primacy Model (Page & Norris, 2009) as it focuses on lower-level features of words influencing Hebb learning, rather than high-level features such as semantics.

However, it remains possible that semantic information could still influence Hebb learning, albeit to a smaller magnitude than metrical information and motor responses. In this sense, it might not be possible for semantic information to produce a detectable Hebb repetition effect alone, however, could still provide an additive benefit to Hebb learning. In Experiment 2, we investigate whether once a Hebb sequence is acquired, semantic order knowledge can be transferred to novel sequences.

3. Experiment 2

Experiment 1 failed to show evidence that the HRE can be found when only the semantic pattern of lists is repeated. Unlike the typical Hebb repetition procedure, the 'Hebb' lists used unique items with repeated semantic patterns. This suggests that semantic information is not acquired alongside item and order information in long-term serial order learning (i.e., the HRE). However, whilst semantic repetition may be insufficiently powerful to enable the HRE, it is possible that once a Hebb sequence has been learnt that information can be used to help recall of related lists. This is directly tested in Experiment 2 where following conventional acquisition of a Hebb list; it is tested whether the reduction in recall errors from the acquired Hebb list can also reduce recall errors in novel item lists that follow the same semantic pattern.

Experiment 2 will initially employ a typical Hebb repetition learning procedure, where participants will undertake 7 epochs of experimental trials. Each epoch will comprise a repeated Hebb sequence and two filler (non-repeated) trials. The Hebb sequence will be the same items presented in the same order, whereas the filler trials will be constructed from different items to those used for the Hebb trials (i.e., no stimulus overlap) and are presented in a novel order for each epoch. It is predicted that a HRE effect should be shown for initial learning of the repeated sequence (see Page et al., 2013 and Johnson et al., 2017 for precedents). However, the key comparison for Experiment 2 concerns whether learning of the Hebb sequence confers a recall advantage to a transfer list that has unique items but share the same semantic pattern of items. Specifically, participants will be tested on recall errors for lists comprising of novel items following the learnt semantic pattern, or novel items following a novel semantic pattern. If learning of the semantic pattern from an acquired Hebb list can be transferred to subsequent lists, it is predicted that there will be significantly less recall errors for novel item lists that follow the same semantic pattern as the learnt Hebb list, relative to novel item lists that follow a novel semantic pattern. Such a finding would indicate that a component of that which is learnt for the Hebb list was the semantic order. Experiment 2 therefore has two hypotheses. Firstly, there will be a significant HRE shown through a greater reduction in errors made across repetitions of Hebb lists, relative to filler lists (i.e., a classic HRE). Secondly, novel (but semantically related) Hebb lists will show significantly fewer errors compared to novel filler lists.

3.1. Method

3.1.1. Design

A repeated measures multifactorial design (4x2) was used, where the first factor was trial type (filler, Hebb, transfer filler, and transfer Hebb) and the second factor was position of the 'transfer Hebb' lists (immediately after the final Hebb list or after the final Hebb list and two intervening transfer filler lists). The dependent variable was serial reconstruction errors. As stated in the pre-registration (https://osf.io/35ncw) an additional piece of analysis was conducted with a repeated measure independent variable of trial type (Hebb and filler) and a dependent variable of learning gradient.

There were two experimental blocks, where each block comprised 27 trials. The first 21 trials were divided into seven 3-trial experimental epochs, where each epoch contained two filler trials followed by a Hebb trial. The Hebb trials were comprised of the same six category names in the same serial order. The two filler trials were constructed from two different sets of 6- category words but the order of these words was different across each epoch. The next two trials were non-analysed familiarisation trials, included to enable participants to adjust to the change in stimuli. These familiarisation trials comprised of exemplars that belonged to the six category words used as the Hebb list in the initial learning block. These were arranged in a novel order so not to match the Hebb semantic pattern. The final four trials in each block were the two ‘novel Hebb’ and two ‘novel filler’ trials. The novel Hebb lists were constructed from exemplars of the six category names used in the original Hebb list, with those exemplars presented in the same order as their related category labels. The two novel Hebb lists used a different set of exemplars. The novel filler lists were constructed from the same exemplars as used for the novel Hebb but were presented in a unique order. In one block, the novel filler trials preceded the novel Hebb trials and in the second block, the novel Hebb trials preceded the novel filler trials. The order of blocks was counterbalanced across participants.

3.1.2. Participants

A power simulation was conducted on R (https://www.r-project.org/) using the Superpower package (Lakens & Caldwell, 2020). Here we powered for obtaining the conventional HRE in the first 21 trials of each block. The power calculation was based upon the raw data from Page et al. (2013, Experiment 2) and used data from the no stimulus overlap condition in which the repeated Hebb sequence occurred every third trial (the closest analogue to the present design). The power simulation was based upon having a 0.9 probability of obtaining a significant difference ($\alpha = .05$; with repeated measures ANOVA) between the gradient of improvement for the Hebb and filler conditions (i.e., detection of the classic Hebb repetition effect). The simulation yielded a sample size of 14, however, the present study recruited 75 participants. The higher recruitment target was chosen because the present experiment compares the difference between novel Hebb (i.e., novel list items that follow the repeated semantic pattern) and novel filler (novel list items that also have a novel semantic pattern), as well as analyse the typical HRE from the initial learning block. The former comparison is something of an unknown but would likely produce a smaller effect than the typical HRE, therefore a larger sample was deemed necessary. Furthermore, given the non-significant findings of Experiment 1, and the novel nature of the research, a larger sample size was chosen to maximise the chances of finding significant effects, if they are to be found.

Participants were recruited through the online recruitment website ‘Prolific’ (www.prolific.co) and were paid £3.75 as compensation. All participants were English-speaking and aged 18-50 years old. Overall, 119 (*M* = 27.03, *SD* = 7.33, aged 18-48, 46 females, 1 ‘Prefer Not to Say’) were recruited. Due to the pre-determined performance criteria (see ‘3.1.5 Data Analysis’), 44 participants were excluded from the analysis. The final sample, therefore, consisted of 75 participants (*M* = 26.84, *SD* = 7.07, aged 18-45, 32 females). The age of participants was not normally distributed ($W = 0.91$, $p < .001$). Although there were no age-related outliers there was still a slight skew towards a younger adult age.

3.1.3. Materials

The experiment building software ‘Gorilla’ (build 20201002, https://gorilla.sc/) was used to create and run the present study (Anwyl-Irvine et al., 2020). A total of 108 words were used as stimuli, 36 category words and 72 exemplars (two per category). All words were between 3-8 letters and either one or two syllables. These words were selected from a pilot study in which participants (n=50) chose the best two exemplar words for a category. This was repeated for 36 categories (see Appendix B1 and B2). Words were spoken by a female voice, recorded and edited using the application ‘Audacity’ (version 2.4.2, https://www.audacityteam.org/) to form 500ms sound envelopes.

Category words were split into six sets of six words. Sets 1-3 were used in Block 1 and 4-6 in Block 2. Three versions of the study were created so that each set appeared as the Hebb list once and as a filler list twice. An additional three versions were created in which the position of the transfer Hebb and transfer filler lists were changed. For example, version 1a used set 1 as the Hebb list in Block 1. After the final Hebb list was presented, participants saw two familiarisation trials (non-analysed) two novel Hebb lists, then two novel filler lists. In Block 2, using set 4, the order of the novel lists was reversed. Version 1b used the same sets of words but the order of the novel lists was reversed such that the novel Hebb lists appeared first for set 4 and second for set 1. The presentation of each block was also randomised within each version.

3.1.4. Procedure

The present study was completed online. After obtaining informed consent, participants were asked for their age and gender as demographic information. They completed 50 experimental trials (plus four non-analysed familiarisation trials, see Figure 4 for example lists), split into two blocks, in which each trial comprised the auditory sequential presentation of six category words. Within each block, the first 21 trials comprised the seven 3-trial epochs, where each epoch included two filler trials and one Hebb trial. This was followed by the two (non-analysed) familiarisation trials, and then finally the two novel filler and two novel Hebb trials (the order of which were counterbalanced). Each trial began with a 3000ms interval, followed by each word being presented within a 500ms sound envelope proceeded by a 500ms inter-stimulus interval. The test phase followed a 500ms retention interval in which all six words were presented simultaneously on the screen in a randomised order. Participants were required to click on each word in the order of original presentation. Words could only be selected once and disappeared when selected. There was no time limit for reconstructing the list. Trials were self-paced. On completion of the task, participants were asked three questions regarding the lists they had just learnt. Firstly, participants were asked whether they noticed anything "special" about the lists they had learnt. Secondly, they were asked whether they noticed any repetition in the lists. Thirdly, participants were asked to elaborate on what they thought had been repeated. Participants typed their responses to questions one and three, and a button response (yes or no) was used for question two. The study took approximately 30 minutes.

Hebb	Medic	Insect	Sport	Tree	Dairy	Metal
Filler	Flower	Planet	Colour	Music	Clothing	Snake
Filler	Season	Fruit	Drink	Dance	Shape	Curry
Novel Hebb	Doctor	Earwig	Football	Oak	Cheese	Steel
Novel Filler	Rugby	Copper	Nurse	Beetle	Pine	Milk

Figure 4. Experiment 2 list examples.

3.1.5. Data Analysis

There were two pre-registered performance-based exclusion criteria (https://osf.io/35ncw). First, if the overall average serial recall error for filler lists and the first Hebb list (i.e., all the novel unrepeated lists) was 15% or below, the participant was excluded, as near ceiling performance would prevent the detection of any Hebb repetition learning. Second, if the overall average serial recall error for the same novel unrepeated lists was 65% or above, the participant was excluded. This is because 66% error could be obtained by remembering one item and guessing the remaining five items. A more conservative exclusion criteria was adopted for the next two experiments as it was thought that the exclusion criteria from Experiment 1 was too restrictive.

The initial learning of the semantic Hebb sequence was analysed using generalised linear mixed effect models (GLME) in R Studio (R Studio Team, 2021) using the lme4 package (version 1.1.26). The model used the fixed effects of trial type (Hebb and the second filler trial of each epoch) and experimental epoch (1-7). A random intercepts model was used as a function of stimulus item and participant. The dependent variable was serial recall error for the Hebb and filler sequences.

The transfer of the learnt semantic Hebb sequence was analysed using generalised linear mixed effect models (GLME) in R Studio (R Studio Team, 2021) using the lme4 package (version 1.1.26). Following Barr et al's. (2013) procedure, the largest random effect structure that converged would be reported. The model used the fixed effects of trial type (novel Hebb lists and novel filler lists) and novel filler position (before or after the two novel Hebb sequences). A random slopes model was used as a function of stimulus item and participant for the fixed factor of trial type. The dependent variable was serial recall error for the novel Hebb and novel filler sequences.

Analysis was also conducted using General Linear Models (i.e., *t*-tests etc.) alongside the Generalised Linear Mixed Effects Model analysis (see section '2.1.5. Data Analysis' for justification). Bayes factors were also reported alongside the General Linear Model analysis to provide an additional measure of effect strength, using the typical cut offs for substantial evidence of the alternative hypothesis ($BF_{10} = 3$) or null hypothesis ($BF_{01} = 0.33$, Bolstad & Curran, 2016).

3.2. Results

3.2.1. Descriptive Data

A serial position curve, seen in Figure 5, shows the typical curve expected for auditorily presented lists, with a strong primacy and slight recency effect for both the Hebb and filler functions (Drewnowski & Murdock, 1980). There is strong primacy as serial position one and two show recall advantages, however, slight recency because there is only a recall advantage shown at serial position six.

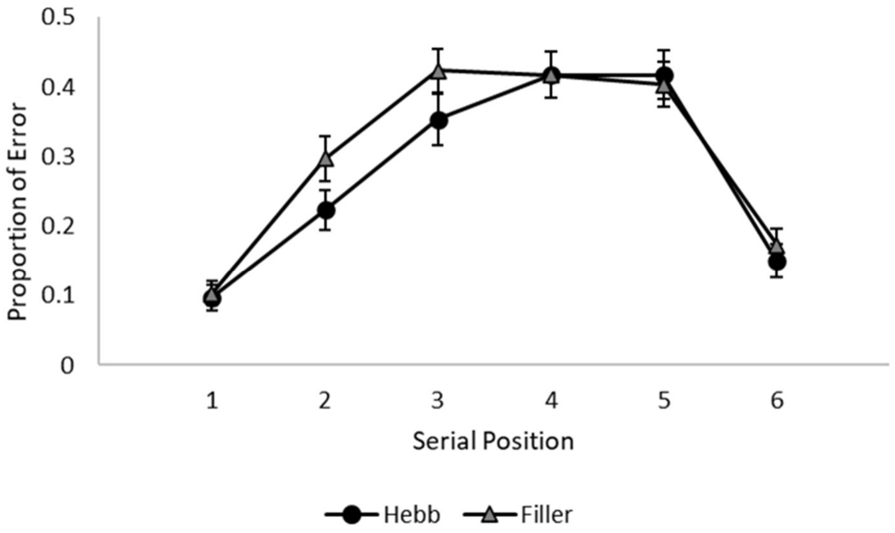


Figure 5. Mean proportion of errors made for each serial position of Hebb and filler lists. Data is taken from the initial learning blocks (trials 1-21) collapsed across both learning blocks. Error bars represent standard error.

Figure 6 illustrates the learning gradients across experimental epoch for Hebb lists and the second filler of each epoch. A steeper learning gradient can be seen for Hebb lists compared to filler lists, taken together with larger gaps between Hebb and filler error bars in the latter experimental epochs. This suggests a possible interaction between Trial Type and Experimental Epoch wherein a reduction in recall errors is seen across the duration of the experiment, only for Hebb lists (i.e., a typical HRE).

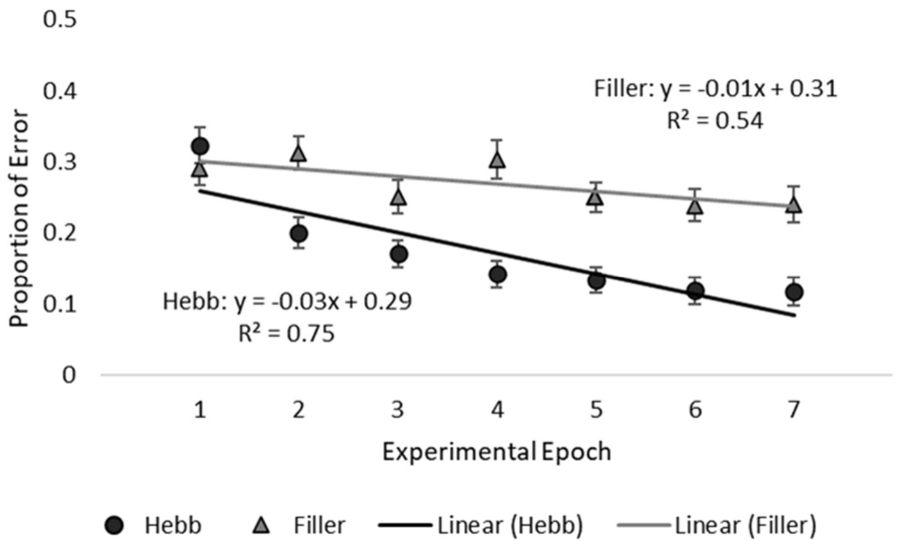


Figure 6. Mean proportion of errors made for Hebb and filler lists according to number of repetitions (experimental epoch). Gradient equation and R-squared values are noted for Hebb and filler lists. Error bars represent standard error.

Figure 7 illustrates the average error rate for novel list items that follow the same Hebb pattern from the learning block, and novel list items following a novel 'filler' pattern. There appears to be some overlap between error bars for novel Hebb and novel filler lists. This suggests that there is no significant difference in recall error for novel item lists that follow the same semantic pattern as the Hebb list (novel Hebb) and novel item lists following a novel semantic pattern (novel fillers).

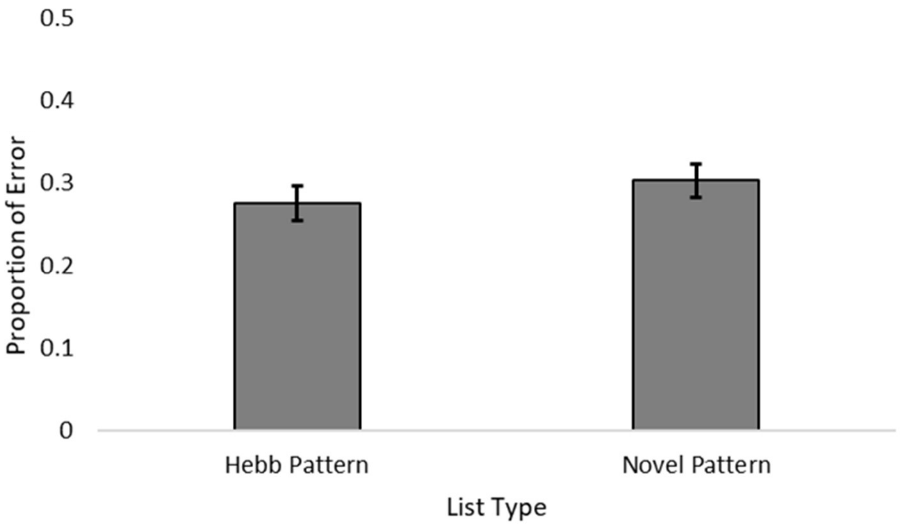


Figure 7. Mean proportion of errors made on novel Hebb (novel items following the previously learnt semantic Hebb pattern) and novel filler (novel items following a novel semantic pattern) lists. Error bars represent standard error.

Descriptive analysis was conducted regarding the awareness questions at the end of the study. When asked whether they noticed anything being repeated, 72 of the 75 participants responded ‘yes’. Although awareness of repetition does not influence typical HREs (Hebb, 1961; McKelvie, 1987), when asked to explain what had been repeated, only 26 participants noted that certain lists were repeated in the exact same order. Most participants noted more generally that certain words were repeated throughout the experiment but did not specify that the order in which words were presented were repeated.

3.2.2. Generalised Linear Mixed Effects Model

3.2.2.1. Initial Learning

Evidence for initial Hebb repetition learning was assessed via a GLME with the fixed factors trial type (Hebb and filler) and experimental epoch (1-7). The GLME model is reported in Table 2 and includes estimates, standard errors, z-values, and p values. The fixed effect of trial type was non-significant. The fixed effect of experimental epoch was significant suggesting general practice-based improvements in serial recall. Importantly, the interaction between trial type and experimental epoch was significant, suggesting a gradual improvement in serial recall as the task progresses but to a greater extent for the Hebb trials (see Figure 6).

Table 2.

Coefficients from the GLME analysis

	Estimate	Std. error	z-value	p-value
Intercept	-0.86	0.12	-7.48	< .01
Trial Type	-0.04	0.10	-0.37	.71
Experimental Epoch	-0.06	0.01	-3.99	< .01
Trial Type*Experimental Epoch	-0.17	0.02	-7.19	< .01

3.2.2.2. Transfer of Semantic Pattern

The extent to which acquired semantic knowledge from the learnt Hebb sequence could be transferred to a semantically related sequence was examined via a GLME with the fixed effects trial type (novel Hebb and novel filler) and filler position (novel Hebb before or after the novel filler trials). The GLME model is reported in Table 3 and includes estimates, standard errors, z-values, and p values. Importantly there was a non-significant fixed effect of trial type, indicating that the difference between the serial recall errors for novel Hebb lists and novel filler lists did not significantly differ. There was no overall effect of the novel filler position on errors and nor did this interact with trial type. The analysis indicates no evidence of serial recall semantic transfer from the repeated Hebb sequence.

Table 3.

Coefficients from the GLME analysis

	Estimate	Std. error	z-value	p-value
Intercept	-1.14	0.17	-6.66	< .01
Trial Type	-0.31	0.26	-1.18	.24
Filler Position	0.08	0.11	0.66	.51
Trial Type*Filler Position	-0.01	0.17	-0.05	.96

3.2.3. General Linear Model

3.2.3.1. Initial Learning

Both Hebb (W = .99, $p = .699$) and filler (W = .98, $p = .398$) lists were normally distributed. A repeated measures t-test found that the error gradient for Hebb lists ($M = -.03$, $SE < .01$) was significantly steeper than the gradient for Filler lists ($M = -.01$, $SE < .01$), $t(74) = 3.37$, $p = .001$, $d = 0.39$, $BF_{10} = 41.98$ (see Figure 6).

3.2.3.2. Transfer of Semantic Pattern

The assumption of normality was violated across all conditions (all W > .91, all $p < .001$). Given a non-parametric alternative does not exist, a 2x2 repeated measures ANOVA was still conducted with the note that results should be treated with caution due to non-normally distributed data. The main effects of Trial Type, $F(1, 74) = 1.72$, $MSE = 0.03$, $p = .193$, $\eta_p^2 = 0.02$, $BF_{10} = 0.27$, and Position, $F(1, 74) = 0.31$, $MSE = 0.04$, $p = .577$, $\eta_p^2 < .01$, $BF_{10} = 0.15$, were not significant. The interaction between Trial Type and Position was not significant, $F(1, 74) = 0.05$, $MSE = 0.03$, $p = .830$, $\eta_p^2 < .01$, $BF_{10} = 0.01$.

3.3. Discussion

The aim of Experiment 2 was to further investigate if there is a Hebb effect for semantic information by establishing whether once a typical Hebb list has been learnt, can the learnt sequence knowledge be transferred to a novel item list that follows the same semantic pattern. This would point towards semantic information being acquired alongside item and order information. However, contrary to that prediction, no evidence of transfer of ordered semantic learning was found.

Results from both the Generalised Linear Mixed effects model and General Linear model analyses showed a significant Hebb repetition effect in the initial learning stage, with a reduction in recall error seen only for Hebb lists, and no reduction in recall errors seen for filler lists. This finding is consistent with Hebb repetition literature across multiple stimulus types (Hebb, 1961; Johnson et al., 2013; Johnson et al., 2016; Page et al., 2013) providing additional evidence supporting the already robust effect. It is also one of, if not, the first instance of a conventional HRE using online testing, suggesting it is an appropriate method of testing in future experiments if in-person testing is not available. There was, however, no significant difference in recall error for novel item lists following the same semantic pattern as the previous Hebb list (i.e., recall for the transfer Hebb list), compared to novel item lists following a novel semantic pattern. This suggests that the recall advantage seen for a repeated list does not transfer to other lists that follow the same semantic pattern. Furthermore, there was no significant effect of list position, meaning that recall errors for both trial types were unaffected by whether they were presented immediately following learning or after the other trial types (i.e., no recall difference for Hebb then filler lists or filler then Hebb lists).

Experiment 1 definitively demonstrated that a HRE was not produced following repetition of the same semantic sequence. Building on this finding, Experiment 2 has shown that that even when a given list has been learnt (evidenced through the typical Hebb repetition effect seen in the initial learning), there is no transfer of the recall advantage to novel item lists that follow the same semantic pattern. This provides further evidence suggesting that information regarding the semantic pattern of a list is not acquired alongside item and order information during Hebb repetition learning. However, it is important to reiterate that semantic information does have an additive benefit to ISR (Poirier & Saint-Aubin, 1995; Saint-Aubin & Poirier 1999; Saint-Aubin et al., 2005).

One might speculate that in Experiments 1 and 2 the semantic categories were not apparent to the participants and therefore were not encoded as part of the sequence. Although awareness of repetition is not required in conventional HREs (Hebb, 1961; McKelvie, 1987), awareness of exemplars belonging to a given category would likely be required to elicit a semantic HRE (see section '5. General Discussion' for additional explanation). Therefore, in a final study we sought to increase the salience of the semantic categories to ensure participants were aware of exemplars belonging to specific categories and test whether we could encourage semantic learning for the repeated Hebb sequences.

4. Experiment 3

Experiment 2 replicated the classic HRE, with a significantly steeper learning gradient for the Hebb list, relative to filler lists. There was no evidence, however, that learnt sequential knowledge of the semantic pattern was transferred to a novel list. This suggests that even when item and order information about a list is well-established, the semantic information does not transfer to other lists. Taken together, Experiments 1 and 2 suggest that semantic information might not be acquired alongside item and order information during Hebb repetition learning. Whilst the preceding experiments suggest that semantic information is not automatically acquired as part of the HRE, Experiment 3 tests whether semantic learning can be artificially promoted via the learning procedure. Here, participants are presented the Hebb list at encoding using the typical procedure (i.e., they receive the same items in the same order of presentation). Hebb lists are comprised of category words such as 'Animal', 'Flower' or 'Colour'. At test, however, participants are presented with items semantically related to the list (i.e., exemplars of category words such as 'Dog', 'Daisy' or 'Blue') and are required to reconstruct the sequence by selecting items in the same semantic order as the preceding list (SOR). Each repetition of the Hebb list will have a novel set of exemplars for participants to use to reconstruct the semantic category pattern. Filler lists, at encoding, are also category labels, with exemplars being used at test to reconstruct the list (see Figure 8 in 'Section 4.1.3. Materials' for example trial and Figure 9 in 'Section 4.1.4. Procedure' for example Hebb and filler lists). Being tested with exemplars of the categories displayed at encoding was hoped to increase the saliency of the semantic information. If a HRE is present, one explanation might be that participants are learning a semantic pattern. Experiment 3 therefore has one hypothesis that there will be a significant HRE evidenced through a greater reduction in errors made across the experiment for Hebb lists, relative to filler lists.

4.1. Method

4.1.1. Design

A repeated measures multifactorial (2x7) design was used. The first factor was trial type (Hebb or filler), with Hebb lists being repeated and filler lists being novel. The second factor was experimental epoch (epoch 1-7). Each epoch comprised 3 trials where the first two 'filler' trials were comprised of a different set of items (with the order of this set unique at each epoch) and the third trial was a repeated Hebb sequence. The dependent variable was serial reconstruction errors.

4.1.2. Participants

A power analysis was conducted using the 'R' software (https://www.r-project.org/) and the 'simr' package. The data used was taken from the learning block from Experiment 2. The power analysis was conducted on the interaction term, as this shows the Hebb effect itself. The unstandardised effect size for the interaction from the Experiment 2 data was 0.08 however, the effect size used for the power analysis was halved in this calculation as the present task has the added difficulty of reconstructing lists using exemplars, instead of reconstructing lists using the category items presented during encoding. Therefore, it is likely this study will elicit a smaller Hebb effect than a traditional Hebb paradigm. The analysis revealed that with a desired power of 80%, an alpha level of .05 and an effect size of 0.04, 60 participants would need to be recruited.

Participants were recruited from the Psychology department at Bournemouth University, via the Sona participation system (https://www.sona-systems.com). Participants were compensated with course credits. Overall, 74 (*M* = 21.89, *SD* = 6.70, aged 18-49, 67 females) were recruited, however, due to a pre-determined performance criterion (see '4.1.5 Data Analysis'), 14 participants were excluded from the analysis. The final sample, therefore, consistent of 60 participants (*M* = 22.42, *SD* = 7.34, aged 18-49, 54 females). The age of participants was not normally distributed ($W = 0.51$, $p < .001$) with seven age-related outliers and a skew towards a younger adult age.

4.1.3. Materials

The experiment building software ‘Gorilla’ (build 20201002, https://gorilla.sc/) was used to create and run the present study (Anwyl-Irvine et al., 2020). 144 English words, ranging from 3-9 letters and 1-2 syllables, were used as stimuli. Of that stimulus set, 18 were category words (e.g., Fruit, Sport etc.). These words were randomly split into three sets of six words which were used to create the Hebb and filler lists. For every category word, seven exemplar words (126 in total) were generated so that for every epoch, a new set of exemplar words were presented during the recall phase. Three different versions of the task were created wherein a different Hebb sequence was used for each version. Participants were randomly assigned to one of the three versions.

In the previous experiments the exemplars were chosen based on pre-studies, in which participants chose items they thought best represented each category. The exemplars that were most frequently chosen were used in the main experiment to ensure they were collectively considered as ‘good’ examples of each category. This decision was made because in the first two experiments, participants were not made aware of the link between category words and exemplars. Participants had to detect the link themselves, therefore, it was important for exemplars to be ‘good’ examples of said categories, to help participants detect the presence of said categories themselves. For example, without prior knowledge, it would be easier to detect the category ‘Body Part’ with exemplars such as ‘Nose’, Knee’ and ‘Foot’ compared to ‘Limb, ‘Palm’ and ‘Back’. In this experiment, however, participants were told to explicitly listen to category words and link them to the respective exemplars given during recall. Therefore, participants were not required to detect category membership themselves. Subsequently, it was not necessary to conduct a pre-study for this experiment to determine the best exemplars for each category as even a less typical exemplar such as ‘Limb’ would be an obvious member of the category ‘Body Part’ due to participants knowledge of the categories.

In a conventional Hebb repetition procedure, the items presented during the encoding phase of each trial are tested at recall. However, this experiment employed a novel version of SOR, wherein the items re-presented at test differed to those at encoding. Specifically, exemplars of the category items shown at encoding were shown at test (see Figure 8 for example trial). That is, whilst the test items were different to the learning items, they were semantically related. For each trial a different set of exemplars were used at test (even though the same category items were used at encoding).

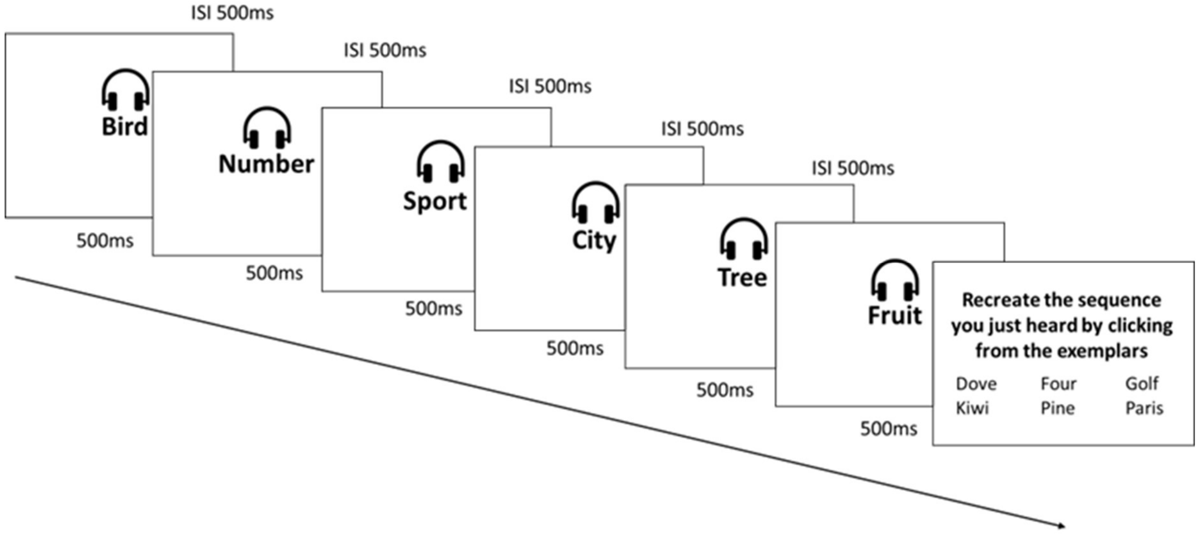
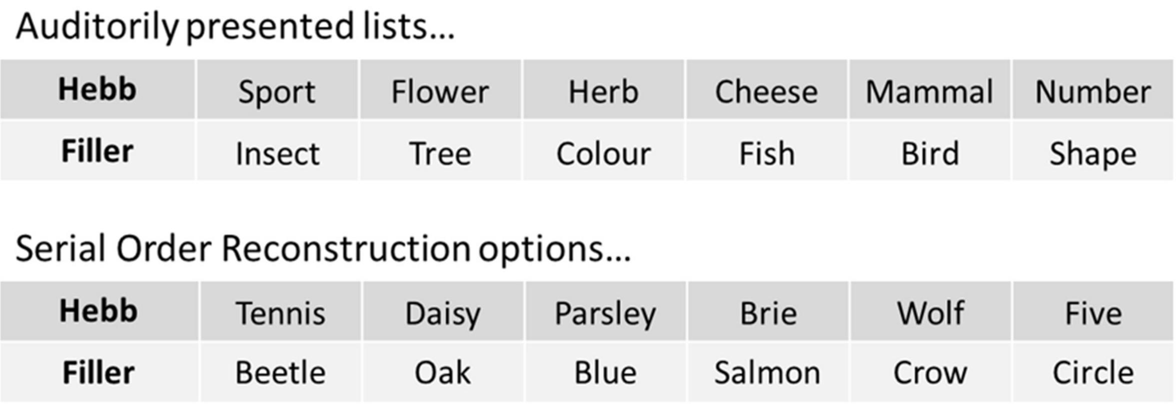


Figure 8. Example trial procedure, showing response exemplars for the Hebb list Bird, Number, Sport, City, Tree, Fruit, including audio durations and Inter-Stimulus Intervals (ISI). Note that Exemplars are not visually presented during encoding and the order of the test words was randomised across trials and participants. At encoding, words were presented

auditorily. Eighteen category words were recorded. These words were recorded by a female voice and each word was cut into 500ms audio files using the application 'Audacity' (version 2.4.2, https://www.audacityteam.org/).

4.1.4. Procedure

The study was conducted online. Participants provided informed consent and then, following additional task instructions, completed five practice trials. Each experimental trial consisted of participants listening to six category words, with a 500ms inter-stimulus interval between the 500ms audio files. At test, six exemplar words were presented in two rows of three boxes, the order of which were randomly presented. To reconstruct the serial order, participants were instructed to click the exemplars that matched the category words in the order of original presentation. Reconstruction time was unconstrained. Unbeknownst to the participants, every third trial the exact category list was repeated. There were 21 experimental trials (see Figure 9 for list examples), with an additional 3000ms in between each trial. Following the 21 trials, participants were asked three questions: first, whether they noticed anything special about the lists they learnt; second, whether they noticed any repetitions; and third, what those repetitions were. Overall, the study took approximately 15 minutes to complete.



Auditorily presented lists...

Hebb	Sport	Flower	Herb	Cheese	Mammal	Number
Filler	Insect	Tree	Colour	Fish	Bird	Shape

Serial Order Reconstruction options...

Hebb	Tennis	Daisy	Parsley	Brie	Wolf	Five
Filler	Beetle	Oak	Blue	Salmon	Crow	Circle

Figure 9. Experiment 3 list examples.

4.1.5. Data Analysis

The same pre-registered performance-based exclusion criteria as Experiment 2 were used for Experiment 3 (see pre-registration details, https://osf.io/wykrh). Participants were excluded if overall average serial recall error for filler lists and the first Hebb list (i.e., all the novel unrepeated lists) was 15% or below, and 65% or above (see section '3.1.5. Data Analysis' for justification).

We analysed learning of the Hebb sequence using generalised linear mixed effect models (GLME) in R Studio (R Studio Team, 2021) using the lme4 package (version 1.1.26,). Following Barr et al's. (2013) procedure, the largest random effect structure that converged would be reported. The model used the fixed effects of trial type (Hebb and the second filler trial of each epoch) and experimental epoch (1-7). A random intercepts model was used as a function of stimulus item and participant. The dependent variable was serial recall error.

Analysis was also conducted using General Linear Models (i.e., *t*-tests etc.) alongside the Generalised Linear Mixed Effects Model analysis (see section '2.1.5. Data Analysis' for justification). Bayes factors were also reported alongside the General Linear Model analysis to provide an additional measure of effect strength, using the typical cut offs for substantial evidence of the alternative hypothesis (BF10 = 3) or null hypothesis (BF01 = 0.33, Bolstad & Curran, 2016).

4.2. Results

4.2.1. Descriptive Data

Figure 10 shows the serial position curves for the Hebb and filler trials and demonstrates the canonical serial position curve expected for ISR of auditorily presented lists, with a strong primacy and a moderate recency effect (Drewnowski & Murdock, 1980).

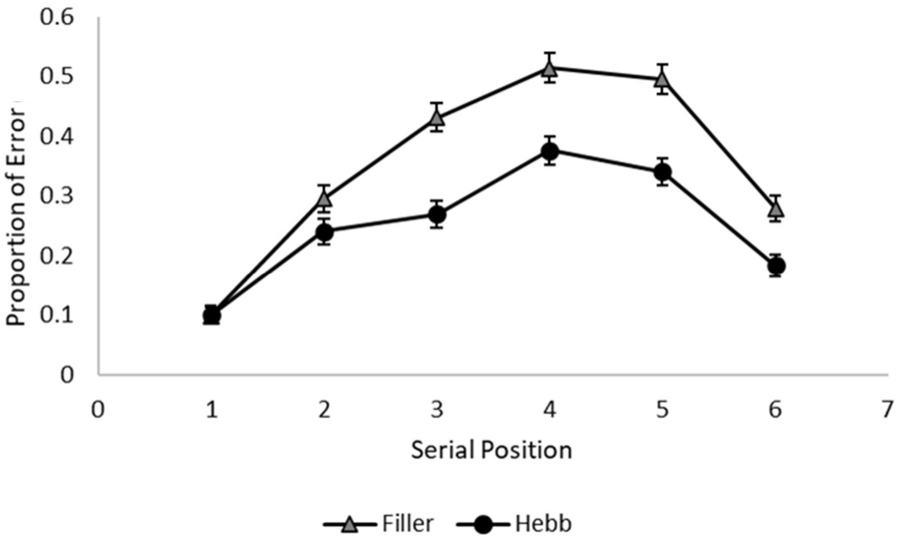


Figure 10. Mean proportion of errors made for each serial position of Hebb and filler lists. Error bars represent standard error.

Figure 11 illustrates the learning gradients for Hebb and filler lists across experimental epoch. There is a steeper learning gradient (i.e., a negative error gradient) for the Hebb list compared to the filler list, as well as generally larger gaps between filler and Hebb error bars in latter experimental epochs (however, note an increase in error for Hebb lists at serial position 7). Taken together, Figure 11 shows a possible interaction between Trial Type and Experimental Epoch wherein a reduction in recall error is found across the duration of the experiment for Hebb lists only (i.e., a typical HRE).

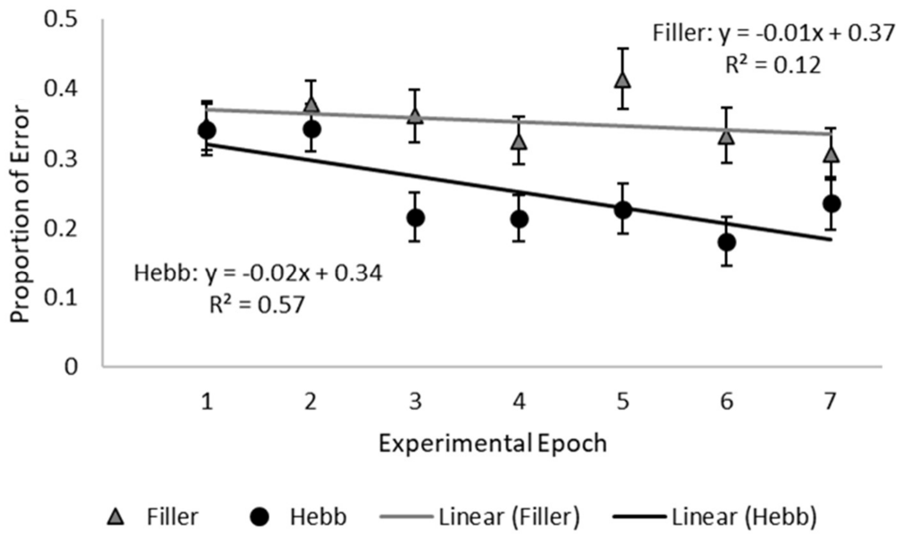


Figure 11. Mean proportion of errors made for Hebb and filler lists according to number of repetitions (experimental epoch). Gradient equation and R-squared values are noted for Hebb and filler lists. Error bars represent standard error.

Descriptive analysis was conducted regarding the awareness questions at the end of the study. When asked whether they noticed anything being repeated, 49 of the 60 participants responded 'yes'. However, when asked to explain what was repeated only 20 participants noted that certain lists were repeated in the exact same order. Most participants noted that the same category labels appeared in many trials but did not specify that these were in a repeated order.

4.2.2. Generalised Linear Mixed Effects Model

Hebb repetition learning was examined via a GLME with the fixed effects of trial type (Hebb and filler) and experimental epoch (1-7). The GLME model is reported in Table 4 and includes estimates, standard errors, z-values, and p values. Both fixed effects of trial type and experimental epoch were non-significant. Importantly, the interaction between trial type and experimental epoch was significant, suggesting a gradual improvement in serial recall as the task progresses but only for the Hebb trials (see Figure 11).

Table 4.

Coefficients from the GLME analysis

	Estimate	Std. error	z-value	p-value
Intercept	-0.58	0.17	-3.31	< .01
Trial Type	-0.13	0.15	-0.88	.38
Experimental Epoch	-0.03	0.02	-1.31	.19
Trial Type*Filler Position	-0.01	0.03	-3.26	< .01

4.2.3. General Linear Model

Consistent with Experiment 2 and previous studies investigating the Hebb repetition effect, we conducted complimentary general linear model analysis on the gradients of improvement. The learning gradient for each list was created by calculating the slope of errors, across epochs, for each participant. The assumption of normality was met for the filler gradient (W = .97, *p* = .079) but was violated for the Hebb gradient (W = .95, *p* = .015). Subsequently, a Wilcoxon Signed Ranks Test was performed. There was a significant difference between filler and Hebb gradient ranks, *Z* = 1158.5, *p* = .037, one-tailed, *r* = .27, BF_{10} = 3.49, with steeper gradients for Hebb lists (*Mdn* = -0.03) compared to filler lists (*Mdn* = -0.01), indicative of a greater reduction in errors for the Hebb sequence as the task progressed.

4.2.4. Additional Exploratory Analysis

Additional exploratory analysis was conducted to compare the initial learning from Experiment 2 and Experiment 3. The procedures of both were almost identical (i.e., seven experimental epochs each containing two fillers and one Hebb), however, differed in terms of whether lists were reconstructed using exact category labels (Experiment 2) or category exemplars (Experiment 3). Note that items and participants also differed between experiments. Two *t*-tests were performed to give an indication as to whether learning rates and overall task difficulty were similar across experiments to help provide an explanation for the significant HRE seen in Experiment 3. Firstly, an Independent Samples *t*-test was conducted comparing Hebb error gradients in Experiment 2 ($M = -0.02$, $SD = 0.05$) and 3 ($M = -0.03$, $SD = 0.03$) and found there was no significant difference in error gradients, $t(133) = 0.93$, $p = .355$, $d = 0.16$. This shows that the HREs seen in Experiment 2 and 3 were highly similar. Secondly, an Independent Samples *t*-test was conducted comparing mean errors for filler lists and found that recall errors were significantly greater for filler lists in Experiment 3 ($M = 0.35$, $SD = 0.03$) compared to Experiment 2 ($M = 0.27$, $SD = 0.03$), $t(12) = 4.90$, $p < .001$, $d = 2.61$. This shows that Experiment 3 was a more challenging task, in general, compared to Experiment 2, which is most likely attributed to reconstructing lists using different, but semantically related items at test.

A third additional exploratory analysis was conducted looking at whether awareness played a role in the significant difference between error gradients seen in Experiment 3. The difference in error gradients was not statistically significant when the 20 participants who reported being aware of a given list being repeated were removed (i.e., only including the 40 participants who were not aware), $Z = 471.5$, $p = .412$, $r = .15$, $BF_{10} = 0.28$. This suggests that participant awareness could have played a role in the Hebb effect seen in Experiment 3 and will be discussed further in the next section.

4.3. Discussion

The aim of Experiment 3 was to explore whether semantic learning could be primed by increasing the salience of category membership during the learning trials. Participants learnt a Hebb list using the typical procedure of repeatedly presenting the same list interspersed with filler lists. The key difference, however, was that at recall participants were shown unique exemplars of the category Hebb list (rather than the category labels used at encoding). This required participants to recall the order of the list, but also transfer that list to fit novel (but semantically related) items.

Results from both the Generalised Linear Mixed Effects model and General Linear Model analysis showed a typical Hebb repetition effect, wherein there was a reduction in errors made in later experimental epochs, only for Hebb lists. This suggests that it is possible for Hebb learning to occur for lists where exact item repetition occurs only during the presentation (or encoding stage) of learning. When recall requires the transfer of category item order to their exemplars, the typical Hebb repetition effect is still found. Overall, the results from Experiment 3 suggest that whilst semantic information can play a role in the Hebb repetition effect when primed, that role is far less pronounced than with item features such as metrical patterns (Paice et al., in preparation) and motor responses (Johnson et al., 2017).

The results of the additional exploratory analysis comparing data from the initial learning blocks in Experiment 2 and Experiment 3 suggest that Hebb learning rates (i.e., error gradients) were highly similar, however, Experiment 3 was, in general, more difficult. The significant difference in filler errors between experiments likely reflects the recall cost associated with converting category labels into their exemplars which increased the task demands in comparison to Experiment 2, wherein lists were reconstructed using the originally presented items. As for the non-significant difference between learning rates, it is possible that participants in Experiment 3 were sub vocally rehearsing the category labels and at test recalled and converted said items to matching exemplars. In this case, participants were still learning and recalling lists in a similar way for Experiment 2 and 3, however, in Experiment 3 there was the additional task demand of converting category labels to exemplars before actually reconstructing the list. This might explain why learning rates were comparable between Experiment 2 and 3, whilst general filler errors were greater for

Experiment 3. This is however speculative and future research would be required to fully understand what caused the HRE seen in Experiment 3 (see ‘5. General Discussion’ for details of future research ideas).

The third additional analysis was conducted on the error gradients for filler and Hebb lists to see whether awareness played a role in the significant difference. When removing participants who specified that they noticed a certain list being repeated multiple times in the exact same order, the significant difference in gradients disappeared. This could suggest that the Hebb effect seen in Experiment 3 was driven by participant awareness of a ‘special’ list that should be remembered. This is, however, a tentative suggestion as removing 20 participants from the analysis would likely make said analysis underpowered.

5. General Discussion

The aim of the three experiments was to investigate whether there exists a Hebb effect for semantic information. Previous studies have investigated semantic effects on short-term list recall (i.e., the semantic similarity effect, Poirier & Saint-Aubin, 1995; Saint-Aubin & Poirier 1999; Saint-Aubin et al., 2005), however, semantic information in long-term learning for lists had not yet been investigated. The present set of studies therefore examined whether general semantic information was acquired through a Hebb repetition learning protocol. In general, findings across the three experiments suggest that semantic information is not acquired alongside item and order information in long-term serial order memory, meaning repetition of semantic patterns alone (i.e., without exact item repetition) cannot produce long-term serial order learning (i.e., the HRE). These findings can be explained by existing models of long-term serial order memory from Page and Norris (2009) and Burgess and Hitch (2006), meaning there is no current need for updating these models to account for semantic HRE findings.

Experiment 1 followed a similar methodological procedure to that of Paice et al. (in preparation) who found that, despite different words being used in each list, repetition of a metrical pattern could produce significant Hebb repetition learning. Similar to that described by Paice et al., the Experiment 1 procedure involved a block of concurrent Hebb trials where the pattern of semantic information was repeated in the same order, but the items used to construct the lists differed in every trial. Unlike the conventional Hebb repetition protocol, there were no intervening filler trials in order to maximise the opportunity of detecting the effect. The comparator filler list trials followed the block of Hebb. However, in contrast to the prediction, in a well-powered study (n = 120), Experiment 1 found that repetition of semantic patterns alone did not produce significant Hebb learning. Evidence in support of the null was found following Bayesian analysis. Given that Experiment 1 followed Paice et al's (in preparation) methodological procedure nearly exactly (who found significant Hebb learning for the repeating metrical patterns), it can be assumed that the non-significant finding in Experiment 1 was not attributed to the blocking of the Hebb trials (i.e., not employing intervening filler trials) or the online testing procedure. This finding is inconsistent with previous HRE literature across a range of stimulus types such as Hebb (1961), Page et al. (2013), Couture and Tremblay (2006) and Horton (2008). The error gradient for Hebb lists in Experiment 1 collapsed across blocks 1-5 was -.001 (.001 if converted to accuracy gradients that are typically used in previous literature). Hebb gradients are typically higher. For example, Page et al. (2013, Experiment 1) reported a mean gradient of .05, Szmalec et al. (2012, Experiment 1) reported a mean gradient of .02 and Johnson et al. (2016) reported a mean gradient of .031 for Hebb lists. In fact, the gradient reported in Experiment 1 is more consistent with filler gradients such as .005 from Page et al. (2013, Experiment 1), .01 from Szmalec et al. (2012, Experiment 1) and .003 from (Johnson et al., 2016). The similarity between the Hebb gradient reported in Experiment 1 and filler gradients from previous literature support the conclusion that no long-term learning of semantic patterns occurred in Experiment 1.

The non-significant findings from Experiment 1 are consistent with the Primacy Model (Page & Norris, 2009) which omit semantic information from the formation of chunks. As a result, a HRE would not be produced by lists that repeat semantic patterns only, instead, exact item repetition would be required. Page and Norris argue that the HRE occurs due to chunking of individual units (such as letter or words etc.) forming a higher-level unit, following repetition. Once a sequence of units has been chunked, more accurate recall would be found as the primacy gradient (i.e., activation grading and suppression mechanisms) prevents recall order errors. This explains why gradual recall improvements are seen in the HRE, because initial presentations of a given list are not chunked. As the list chunk is formed and strengthened (through repetitions) improved recall is shown. According to the Primacy Model, the HRE was not found in Experiment 1 because word units were never repeated in the exact same order, therefore, they were never chunked together to form a higher-level list unit. Subsequently, latter repetitions of the semantic pattern did not produce a reduction in recall errors as words were still being treated as individual word units therefore there was no primacy gradient working to limit errors. If there had been evidence for improved recall of the semantically related Hebb lists, one might suggest it is possible to chunk semantic information in a similar way to item and order information. This would have pointed towards a need to update the Primacy Model to include semantic units within their model. However, because Experiment 1 did not show evidence of semantic information being chunked, it could be argued that Experiment 1 is consistent with the Primacy Model as an explanation of long-term serial order learning (i.e., the HRE).

The non-significant HRE in Experiment 1 can also be explained by Burgess and Hitch's network model (2006). Their model suggests that the HRE occurs due to strengthening associations between items and serial position (i.e., order) allowing for more accurate recall following repetitions of a list. Lists are represented as context sets and when a new list is presented a matching process begins to previous context sets. Previous context sets that reach a threshold of over 0.6 (60% match) remain active, whilst context sets that fall below are inhibited. Burgess and Hitch argue that when a list is repeated there is already an association between item and order from the previous context set. Repeated matching (through list repetition) strengthens the item and order association resulting in a reduction in recall errors for repeated lists compared to non-repeated lists (i.e., the HRE) that are represented as individual context sets. Under this model, a HRE was not found because exact items were not repeated meaning each latter presentation of the Hebb list had a 0% match to the first Hebb list and therefore did not meet the similarity threshold of 0.6 (60%).

Subsequently, all Hebb lists were represented as individual context sets, rather than matching a previous context set, and therefore showed a similar level of recall accuracy to filler lists.

Overall, results from Experiment 1 point towards a view that semantic information is not unitised in a similar way to letters, words, numbers, etc. in the long-term learning of order information, suggesting that ordered semantic information cannot be chunked in a similar way to ordered item information. As a result, repetition of semantic information alone is not able to produce the Hebb repetition effect as seen in previous research (Hebb, 1961; Johnson et al., 2017; Page et al., 2013).

Whilst Experiment 1 demonstrated that repeating ordered semantic information was insufficient to induce a Hebb repetition effect, Experiment 2 tested whether once a Hebb list was learnt, that knowledge could be transferred to a semantically related list. Experiment 2 had two distinct parts (with separate analyses). In the first stage, participants completed a typical Hebb repetition procedure in which they were auditorily presented with six-word lists which they had to reconstruct immediately after the list was presented. Every third trial, the same list was repeated in the same serial order (Hebb list), whilst all other lists were novel filler trials. There were 21 trials overall, with seven 3-trial experimental epochs resulting in seven instances of the following: filler, filler, Hebb trial. This procedure produced a significant Hebb repetition learning effect, evidenced by both the Generalised Linear Mixed Effects and General Linear Model, wherein fewer recall errors were seen in latter experimental epochs for Hebb lists relative to the filler lists. This initial replication of the classic Hebb repetition effect is consistent with past literature using a variety of stimulus types, such as numbers (Hebb, 1961), words (Page et al., 2013), faces (Johnson et al., 2017), odours (Johnson et al., 2013) and touch (Johnson et al., 2016), and shows that the effect can be detected using online testing.

The learning gradients themselves are also consistent with previous literature, albeit the present research analysed errors as opposed to accuracy therefore gradients would be inverted to compare to existing literature that analyses accuracy. Experiment 2 found an error gradient of -.03 for Hebb lists and -.01 for filler lists (i.e., .03 and .01 respectively if looking at accuracy gradients). This is consistent with a number of experiments that followed a similar methodology of Hebb repetitions every third trial with no item overlap between filler and Hebb lists. For example, Page et al. (2013, Experiment 1) administered a Hebb repetition task using lists of seven four-letter words and found a learning gradient of .051 for Hebb lists and .01 for filler lists. Szmalec et al. (2012, Experiment 1) administered a Hebb repetition task using lists of nine syllables and found a learning gradient of .02 for Hebb lists and .005 for filler lists. Johnson et al. (2016) administered a tactile Hebb repetition task using lists of finger touches and found a learning gradient of .031 for Hebb lists and .006 for filler lists. Finally, Johnson et al. (2017, Experiment 1) administered a Hebb repetition task using sequences of five unfamiliar faces and found a learning gradient of .025 for Hebb lists and .006 for filler lists. The above examples show very similar learning gradients for both Hebb and filler lists to what was reported in Experiment 2. Overall, evidence from significant HREs and specific learning gradients across multiple experiments with varying stimulus type suggests that the HRE seen in the learning block of Experiment 2 is consistent with previous Hebb repetition literature. Moreover, this contributes to a broader argument in Hebb repetition literature of whether the effect is purely a phonological, language-based effect. The gradients reported in Experiment 2 (once inverted to reflect accuracy instead of error) reflect a verbal HRE that is very similar to gradients reported by Johnson et al. (2016) and Johnson et al. (2017, Experiment 1) who found a tactile and visual HRE, respectively. This contributes to the argument that the HRE is a general feature of memory and not confined to verbal memory (Couture & Tremblay, 2006; Johnson and Miles, 2019a,b).

Following the learning stage, transfer of semantic pattern learning was then tested. Participants completed six additional trials. The first two were always familiarisation trials which consisted of novel items. These two trials were designed to remove any direct carry-over effects from the final presentation of the Hebb list. Next, participants were presented with two novel Hebb lists (i.e., novel items that followed the same semantic pattern as the learnt Hebb list) and two novel filler lists (i.e., novel items with a novel semantic pattern). Both Generalised Linear Mixed Effects and General Linear Model analyses found there was no significant difference in recall error for novel Hebb and filler lists. Indeed, Bayesian analysis supported the null hypothesis of there being no difference between the transfer Hebb and transfer filler trials. This means that following a well-powered study (n = 75), participants were no better at recalling lists that followed a previously learnt lists semantic pattern than lists following a non-learnt semantic pattern. Building on the first experiment, Experiment 2 shows that once a sequence has been learnt, knowledge cannot be transferred to a semantically related sequence.

The non-significant difference between novel Hebb and filler list recall in Experiment 2 is again consistent with Page and Norris' (2009) Primacy Model as the model does not include semantic information in the formation of chunks. When the word 'dog' is presented, the strongest activation comes from the 'dog' word unit, however, there is also lesser activation of individual letters from the word 'dog' as well as some activation of similar words such as 'dot' or 'doe'. Although semantically related, the word unit 'cat' would not be activated after presentation of 'dog' because it shares none of the same letters. It would only be possible for 'cat' to be activated following presentation of the word 'dog' if semantic information relating to category membership was held within each word unit, which is not currently addressed in the Primacy Model. This explains why there was no reduction in recall errors for novel Hebb lists compared to novel filler lists as semantic information is not held in word or list units; consequently, both novel Hebb and filler units were treated as new unrelated lists.

In contrast, initial learning of the Hebb sequence in Experiment 2 is consistent with Page and Norris' (2009) Primacy Model. Gradual improvement for the Hebb list is explained by the six items in the Hebb list being combined into a new list unit following repetition, so in latter presentations, the new list unit was activated, as opposed to the individual word units. When filler lists, and the first presentation of a Hebb list, are presented, items within the list have not been chucked together. Although primacy gradient mechanisms exist in all units, when items are only activated by individual word units (and not as a chunked list unit), there is no relation between specific items in a list. The primacy gradient for an individual word unit relates to promoting ordered recall of the letters within that word. The primacy gradient mechanism therefore cannot benefit ordered recall of fillers or the initial presentations of Hebb lists. Once a list has been successfully chunked through repetition, the primacy gradient mechanism relates to words within the list and therefore, the primacy gradient aids ordered recall of lists.

Both findings from Experiment 2 could also be explained by an alternative model from Burgess and Hitch (2006). As discussed above, Burgess and Hitch argue that the HRE occurs due to matching previous context sets (i.e., list representations) and with continued matching, stronger associations between items and their serial positions (i.e., list order). This would explain why a significant HRE was found in the initial learning block analysis. The model states that lists that do not match a previous context set to at least a threshold of 0.6 (60%) are inhibited, explaining why there was no apparent transfer of Hebb learning to novel Hebb lists, relative to novel filler lists, because list items did not match the learnt context set (i.e., the Hebb list). Overall, results from Experiment 2 support the conclusion from Experiment 1 that semantic information might not be acquired alongside item and order information in long-term serial order learning. Also, under the Primacy Model's (Page & Norris, 2009) explanation of the Hebb repetition effect, Experiment 2 is also consistent with Experiment 1 in the view that semantic information is not unitised in the same way as information such as letter or words.

Experiments 1 and 2 have shown that repeating semantic patterns cannot produce a Hebb repetition effort nor can a learnt Hebb sequence be transferred to novel semantically related lists. However, it is possible that these null effects were a result of participants being unaware of the semantic relations between list items. Experiment 3 sought to test this explanation by increasing the salience of the semantic categories. Experiment 3 followed a typical Hebb repetition procedure to past studies (Hebb, 1961; Horton, 2008; Johnson et al., 2013), with an exact sequence repetition occurring every third trial. The critical difference, however, is that at test participants were presented with different (albeit semantically related) items to those presented at the encoding stage. Specifically, the to-be-remembered lists were comprised of semantic category names such as 'flower', 'sport', 'animal' etc. At test, participants were presented with exemplars of the categories they had seen, such as, 'daisy', 'tennis', 'dog', etc., which they used to reconstruct the lists. Using exemplars at test was intended to highlight the semantic information in the to-be-remembered lists and therefore repeatedly encode that information when learning the Hebb list. Both Generalised Linear Mixed Effects and General Linear Model analyses found a significant Hebb repetition effect, in that recall errors were significantly reduced across experimental epochs to a greater extent for the Hebb lists.

The significant HRE in Experiment 3 is consistent with previous Hebb literature using verbal stimuli, such as Page et al. (2013) and Szmalec et al. (2012), as well as non-verbal stimuli such as unfamiliar faces (Johnson et al., 2017), visuospatial stimuli (i.e., dots, Couture & Tremblay, 2006), odours (Johnson et al., 2013) and tactile stimuli (i.e., finger touches, Johnson et al., 2016). The specific learning gradients reported in Experiment 3 are also consistent with a number of previous experiments. Experiment 3 reported a median error gradient of -.033 for Hebb lists and -.012 for filler lists (the median was reported due to the data violating the assumption of normality). As with Experiment 2, Experiment 3 analysed error as opposed to accuracy, which is more commonly reported in previous literature but can still be compared by inverting the gradients to .033 and .012 to reflect Hebb and filler learning respectively. Page et al. (2013, Experiment 1) reported a mean learning gradient of .051 for Hebb lists and .01 for filler lists. Szmalec et al. (2012, Experiment 1) reported a mean learning gradient of .02 for Hebb lists and .005 for filler lists. Johnson et al. (2016) reported a mean learning gradient of .031 for Hebb lists and .003 for filler lists. Lastly, Johnson et al. (2017, Experiment 1) reported a mean learning gradient of .025 for Hebb lists and .006 for filler lists. Taken together, evidence of significant HREs and specific examples of learning gradients suggest that the results for Experiment 3 are consistent with previous Hebb repetition literature.

One interpretation of Experiment 3 is that once semantic categories are made more salient to the participants, Hebb repetition learning of repeated semantic information is possible. In this case, the conclusion of Experiment 3 (alongside Experiments 1 and 2) would be that semantic information does impact long-term serial order learning, however, it is such a small effect that it is not detectable unless the semantic aspect is explicitly shown to participants (i.e., by informing participants to convert a presented category list into exemplars). This would prompt a revision of the Primacy Model (Page & Norris, 2009) to include semantic units or the ability to hold semantic information within word units.

Conversely, it is entirely possible that the result of Experiment 3 is not a semantic Hebb effect, but rather participants sub vocally rehearsing the category labels at test and selecting the exemplars that best match the category word from serial position one, two, three etc. Converting the original category labels into exemplars would arguably be a more challenging task compared to a typical Hebb task. Additional exploratory analysis supports this as filler errors were significantly greater in Experiment 3 compared to Experiment 2, suggesting the task in general was more challenging compared to Experiment 2.

It is also worth noting that Hebb repetition learning can occur without recall following repeated presentation of the list only. For example, Oberauer and Meyer, (2009) reported a significant HRE for lists that were presented to participants but not required to be recalled. This effect was similar to the HRE seen when participants were presented with lists and recalled them, as per the typical HRE procedure (see also, Kalm & Norris, 2016). Given that repeated recall at test is not required for the Hebb repetition effect, it is possible that the Hebb repetition effect in Experiment 3 might occur despite the use of novel category exemplars at test. Specifically, participants might be learning the sequence through repeated exposure to the list during encoding. At test, participants might then rely upon sub-vocal rehearsal of the category words and reconstruct the list by recalling categories and matching them to the exemplars. This would explain why there was no significant difference in learning rates between experiments as recall of lists are not required to produce robust HREs. Taken together, the evidence suggesting that Experiment 3 was more difficult than Experiment 2, along with evidence from Kalm and Norris (2016) and Oberauer and Meyer (2009), indicate that it is possible that the result from Experiment 3 were not a result of semantic information influencing long-term serial order learning. A more prosaic, albeit speculative, explanation is that Experiment 3 shows a typical Hebb repetition effect resulting from repeated presentation of a given list during the encoding phase. Assuming the effect occurred purely from repeated presentation (the encoding stage) and had no relevance to semantic patterns, the findings could be explained by the Primacy Model (Page & Norris, 2009) as well as Burgess and Hitch's (2006) model as discussed in relation to Experiment 2's learning block findings.

Although this seems like the more logical explanation for Experiment 3, it is slightly speculative and future research should be conducted to provide a stronger explanation for this experiments results.

The present procedure could be amended to directly test whether the results of Experiment 3 occurred due to semantic information influencing the HRE (as opposed to a HRE from sub-vocal rehearsal of category words at the learning stage). Firstly, Experiment 3 could be repeated with the inclusion of a concurrent articulation manipulation to prevent participants from being able to sub-vocally rehearse the category words. If learning persists under CA, it suggests that participants were not comparing the exemplars at test with the sub-vocally rehearsed category labels presented at learning. Secondly, Experiment 3 could be repeated with the addition of novel Hebb and filler trials (i.e., in Experiment 2), wherein participants are tested on lists with novel items that follow the semantic Hebb pattern of the learnt Hebb list. If superior recall is found for the novel Hebb list (compared to the novel filler list) it would suggest that some learning of the semantic pattern has occurred.

One limitation relating to Experiment 1 and 2 is whether semantic categories were detectable by participants as they were not explicitly told prior to participation that there would be a semantic link between trials. Although previous research has found that the HRE exists regardless of participants awareness that lists have been repeated (Hebb, 1961; McKelvie, 1987), conscious awareness of a semantic link between exemplars and categories would likely be required to elicit a semantic HRE. For example, if a participant is learning the semantic pattern ‘Animal, Music, Place’ and they are unaware that exemplars belong to said categories, a given list would not be considered an exact repetition of the semantic pattern under either Page and Norris (2009) or Burgess and Hitch’s (2006) models. This potential issue was mitigated by including stimuli generation pilot studies for Experiment 1 and 2 (see Appendix A1, A2, B1 and B2) designed to ensure words used as stimuli were considered the most common exemplars by a larger group of people (n = 50). A stimulus generation study was not deemed necessary in Experiment 3 as participants were explicitly told of the link between exemplars and categories, making participants aware of the semantic link. Although these efforts were made to ensure stimuli were common exemplars of categories, evidence from the awareness questions at the end of each experiment suggest that participants may not have detected the categories. For example, only two of the 120 participants in Experiment 1 reported that semantic patterns were repeated, suggesting the vast majority of participants in Experiment 1 were unaware, or chose not to report the link. Whilst more participants from Experiment 2 (n = 26) reported noticing a list was repeated, this does not directly inform us about whether those participants noticed that the novel Hebb trials followed the same semantic pattern (i.e., indicating awareness that a semantic pattern was being repeated as opposed to simply stating that exact items were repeated in the learning blocks). Taken together, it is somewhat unclear whether the non-significant semantic Hebb effect (Experiment 1) and non-significant difference between novel Hebb and novel fillers (Experiment 2) were due to participants being unaware of the semantic link or because the effect genuinely does not exist.

A second limitation of the three experiments is a lack of control of psycholinguistic variables known to impact recall. For example, word frequency (how frequently words are used in written or spoken language, Allen & Hulme, 2006) and concreteness (how concrete or abstract words are, Romani, McAlpine & Martin, 2008). Miller and Roodenrys (2009) found poorer recall accuracy for items with low word frequencies (used less frequently in language) and concreteness (abstract words). It is therefore possible that performance could have been impacted by these variables as it is unknown whether Hebb and filler lists differed in terms of word frequency or concreteness. Future work should endeavour to control for these psycholinguistic variables to reduce potential unexplained variance.

The stimuli generation pilot studies (see Appendices A1 to B2), however, likely controlled for measures of semantic relatedness such as output-dominance. This refers to how related an exemplar is to a category, for example, dog would have a higher output-dominance than deer, for the category mammal (Marsh et al., 2008). The decision to use categories and exemplars, as opposed to other semantically linked items such as synonyms and antonyms (i.e., good-kind, or good-bad, respectively), was based on the constraints of the method used (i.e., SOR of verbal lists). To follow the methodology of Paice et al. (in preparation) closely, lists needed to be constructed wherein six items followed specific semantic patterns with exact items changing across trials. To keep stimuli in the verbal domain, as opposed to pictures used in Page et al. (2006, Experiment 2), patterns of broad verbal categories containing numerous exemplars were deemed an appropriate way to create the Hebb and filler trials across the three experiments. Of course, other ways to operationalise semantic information, such as synonyms and antonyms might yield different findings if used alongside alternative methodological procedures tapping into long-term learning of information. Indeed, ISR and SOR are not the only procedures that can be used to demonstrate Hebb learning. Matching or triad tasks, that are more commonly used in semantic processing literature (Brewer, 1995), could also be used to demonstrate Hebb learning. It is worth noting that different findings could be shown using these types of tasks that rely more on semantic processing, as opposed to the chosen procedure (relying more on articulatory and perceptual processing), which should be considered in future research.

5.1. Conclusion

To conclude, the present research sought to investigate whether a Hebb effect exists for semantic information. Results from the three experiments did not show evidence that Hebb repetition learning occurs with repeated semantic patterns alone despite each sequence being constructed of different items. This indicates that semantic information is not acquired alongside item and order information in the long-term learning of sequences. This means it is unlikely that semantic information is unitised in the same way as information such as phonemes, words or numbers are, meaning that the Primacy Model (Page & Norris, 2009) does not require revision to accommodate semantic learning. Furthermore, results across the three experiments can also be explained by Burgess and Hitch's (2006) model, supporting the conclusion that semantic patterns alone do not produce a HRE and suggesting their model does not require revision to accommodate semantic learning. However, whether there is an additive benefit of semantically related information on Hebb learning remains unclear, leaving scope for future research to continue investigating whether there exists a Hebb effect for semantic information.

References

Allen, R., & Hulme, C. (2006). Speech and language processing mechanisms in verbal serial recall. *Journal of Memory and Language*, *55*(1), 64-88. https://doi.org/10.1016/j.jml.2006.02.002

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods, 52*(1), 388-407. https://doi.org/10.3758/s13428-019-01237-x

Attout, L., Ordonez Magro, L., Szmalec, A., & Majerus, S. (2020). The developmental neural substrates of Hebb repetition learning and their link with reading ability. *Human Brain Mapping*, *41*(14), 3956-3969. https://doi.org/10.1002/hbm.25099

Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, *4*(11), 417-423. https://doi.org/10.1016/S1364-6613(00)01538-2

Baddeley, A. D. (1986). *Working memory*. Oxford, England: Clarendon Press.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. A. Bower (Ed.). *Psychology of Learning and Motivation* (Vol. 8, pp. 47-89). Academic press. https://doi.org/10.1016/S0079-7421(08)60452-1

Baddeley, A., Gathercole, S., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review*, *105*(1), 158-173. https://doi.org10.1037/0033-295x.105.1.158

Baddeley, A., Lewis, V., & Vallar, G. (1984). Exploring the articulatory loop. *The Quarterly Journal of Experimental Psychology Section A*, *36*(2), 233-252. https://doi.org/10.1080/14640748408402157

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255-278. https://doi.org/10.1016/j.jml.2012.11.001

Bogaerts, L., Szmalec, A., De Maeyer, M., Page, M. P. A., & Duyck, W. (2016). The involvement of long-term serial-order memory in reading development: A longitudinal study. *Journal of Experimental Child Psychology, 145*, 139–156. https://doi.org/10.1016/j.jecp.2015. 12.008

Bogaerts, L., Szmalec, A., Hachmann, W. M., Page, M. P. A., & Duyck, W. (2015). Linking memory and language: Evidence for a serial-order learning impairment in dyslexia. *Research in Developmental Disabilities*, *43*, 106–122. https://doi.org/10.1016/j.ridd.2015.06.012

Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H. H., & White, J. S. S. (2009). Generalized linear mixed models: A practical guide for ecology and evolution. *Trends in Ecology & Evolution*, *24*(3), 127-135. https://doi.org/10.1016/j.tree.2008.10.008

Bolstad, W. M., & Curran, J. M. (2016). *Introduction to Bayesian statistics*. New Jersey: John Wiley & Sons.

Botvinick, M. M., & Plaut, D. C. (2006). Short-term memory for serial order: a recurrent neural network model. *Psychological Review*, *113*(2), 201-233. https://psycnet.apa.org/doi/10.1037/0033-295X.113.2.201

Bower, G. H., & Winzenz, D. (1969). Group structure, coding, and memory for digit series. *Journal of Experimental Psychology*, *80*, 1-17. https://psycnet.apa.org/doi/10.1037/h0027249

Brewer, D. D. (1995). Cognitive indicators of knowledge in semantic domains. *Journal of Quantitative Anthropology*, *5*(2), 107-128.

Burgess, N., & Hitch, G. J. (1999). Memory for serial order: A network model of the phonological loop and its timing. *Psychological Review*, *106*(3), 551-581. https://doi.org/10.1037/0033-295X.106.3.551

Burgess, N., & Hitch, G. J. (2006). A revised model of short-term memory and long-term learning of verbal sequences. *Journal of Memory and Language*, *55*(4), 627-652. https://doi.org/10.1016/j.jml.2006.08.005

Couture, M., & Tremblay, S. (2006). Exploring the characteristics of the visuospatial Hebb repetition effect. *Memory & Cognition*, *34*(8), 1720-1729. https://doi.org/10.3758/BF03195933

Cumming, N., Page, M., & Norris, D. (2003). Testing a positional model of the Hebb effect. *Memory*, *11*(1), 43-63. https://doi.org/10.1080/741938175

Cunningham, T. F., Healy, A. F., & Williams, D. M. (1984). Effects of repetition on short-term retention of order information. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10*(4), 575–597. https://doi.org/10.1037/0278-7393.10.4.575

Dell, G. S., & O'Seaghdha, P. G. (1991). Mediated and convergent lexical priming in language production: A comment on Levelt et al (1991). *Psychological Review, 98*(4), 604–614. https://doi.org/10.1037/0033-295X.98.4.604

Di Betta, A. M., & Romani, C. (2006). Lexical learning and dysgraphia in a group of adults with developmental dyslexia. *Cognitive Neuropsychology*, *23*(3), 376-400. https://doi.org/10.1080/02643290442000545

Drewnowski, A., & Murdock, B. B. (1980). The role of auditory features in memory span for words. *Journal of Experimental Psychology: Human Learning and Memory*, *6*(3), 319-332. https://psycnet.apa.org/doi/10.1037/0278-7393.6.3.319

Fastame, M. C., Flude, B., & Hitch, G. J. (2005). How is the serial order of a verbal sequence coded? Some comparisons between models. *Memory, 13*, 247– 258. https://doi.org/10.1080/09658210344000314

Fendrich, D. W., Healy, A. F., & Bourne, L. E. (1991). Long-term repetition effects for motoric and perceptual procedures. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*(1), 137- 151. https://doi.org/10.1037/0278-7393.17.1.137

Guérard, K., Saint-Aubin, J., Boucher, P., & Tremblay, S. (2011). The role of awareness in anticipation and recall performance in the Hebb repetition paradigm: implications for sequence learning. *Memory & Cognition*, *39*(6), 1012-1022. https://doi.org/10.3758/s13421-011-0084-1

Hay, D. C., Smyth, M. M., Hitch, G. J., & Horton, N. J. (2007). Serial position effects in short-term visual memory: A SIMPLE explanation? *Memory & Cognition*, *35*(1), 176-190. https://doi.org/10.3758/BF03195953

Hebb, D. O. (1961). Distinctive features of learning in the higher animal. In J. Delafresnaye (Ed.), *Brain mechanisms and learning* (pp. 37–46). Oxford: Blackwell.

Hebb, D. O. (2005). *The organization of behavior: A neuropsychological theory*. New York: Psychology Press.

Henry, L. A., Poloczek, S., Messer, D. J., Dennan, R., Mattiauda, E., & Danielsson, H. (2022). Hebb repetition learning in adolescents with intellectual disabilities. *Research in developmental disabilities*, *125*, 104219. https://doi.org/10.1016/j.ridd.2022.104219

Hitch, G. J., Flude, B., & Burgess, N. (2009). Slave to the rhythm: Experimental tests of a model for verbal short-term memory and long-term sequence learning. *Journal of Memory and Language*, *61*(1), 97-111. https://doi.org/10.1016/j.jml.2009.02.004

Horton, N., Hay, D. C., & Smyth, M. M. (2008). Hebb repetition effects in visual memory: The roles of verbal rehearsal and distinctiveness. *Quarterly Journal of Experimental Psychology*, *61*(12), 1769-1777. https://doi.org/10.1080%2F17470210802168674

Johnson, A. J., & Miles, C. (2019a). Visual hebb repetition effects: The role of psychological distinctiveness revisited. *Frontiers in Psychology*, *10*, 17. https://doi.org/10.3389/fpsyg.2019.00017

Johnson, A. J., & Miles, C. (2019b). Visual Hebb repetition effects survive changes to both output order and concurrent articulation. *Journal of Cognitive Psychology*, *31*(3), 276-284. https://doi.org/10.1080/20445911.2019.1586715

Johnson, A. J., Cauchi, L., & Miles, C. (2013). Hebbian learning for olfactory sequences. *Quarterly Journal of Experimental Psychology*, *66*(6), 1082-1089 https://doi.org/10.1080/17470218.2012.729068

Johnson, A. J., Dygacz, A., & Miles, C. (2017). Hebb repetition effects for non-verbal visual sequences: Determinants of sequence acquisition. *Memory*, *25*(9), 1279-1293. https://doi.org/10.1080/09658211.2017.1293692

Johnson, A. J., Shaw, J., & Miles, C. (2016). Tactile order memory: evidence for sequence learning phenomena found with other stimulus types. *Journal of Cognitive Psychology*, *28*(6), 718-725. https://doi.org/10.1080/20445911.2016.1186676

Kalm, K., & Norris, D. (2016). Recall is not necessary for verbal sequence learning. *Memory & cognition*, *44*(1), 104-113. https://doi.org/10.3758/s13421-015-0544-0

Lakens, D., & Caldwell, A. R. (2021). Simulation-based power analysis for factorial analysis of variance designs. *Advances in Methods and Practices in Psychological Science*, *4*(1), 2515245920951503. https://doi.org/10.1177%2F2515245920951503

Marsh, J. E., Hughes, R. W., & Jones, D. M. (2008). Auditory distraction in semantic memory: A process-based approach. *Journal of Memory and Language*, *58*(3), 682-700. https://doi.org/10.1016/j.jml.2007.05.002

McKelvie, S. J. (1987). Learning and awareness in the Hebb digits task. *The Journal of General Psychology*, *114*(1), 75-88. https://doi.org/10.1080/00221309.1987.9711057

Melton, A. W. (1967). Relations between short-term memory, long-term memory and learning. In *The organisation of recall: Proceedings of the second international interdisciplinary conference on learning remembering and forgetting* (Vol. 2, pp. 24-62). New York, NY: New York Academy of Sciences.

Miller, L. M., & Roodenrys, S. (2009). The interaction of word frequency and concreteness in immediate serial recall. *Memory & Cognition*, *37*(6), 850-865. https://doi.org/10.3758/MC.37.6.850

Mosse, E. K., & Jarrold, C. (2008). Hebb learning, verbal short-term memory, and the acquisition of phonological forms in children. *Quarterly journal of experimental psychology*, *61*(4), 505-514. https://doi.org/10.1080/17470210701680779

Mosse, E. K., & Jarrold, C. (2010). Searching for the Hebb effect in Down syndrome: evidence for a dissociation between verbal short-term memory and domain-general learning of serial order. *Journal of Intellectual Disability Research*, *54*(4), 295-307. https://doi.org/10.1111/j.1365-2788.2010.01257.x

Norris, D., Page, M. P., & Hall, J. (2018). Learning nonwords: The Hebb repetition effect as a model of word learning. *Memory*, *26*(6), 852-857. https://doi.org/10.1080/09658211.2017.1416639

Oberauer, K., & Meyer, N. (2009). The contributions of encoding, retention, and recall to the Hebb effect. *Memory*, *17*(7), 774-781. https://doi.org/10.1080/09658210903107861

Page, M. P. A., & Norris, D. (2009). A model linking immediate serial recall, the Hebb repetition effect and the learning of phonological word forms. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1536), 3737-3753. https://doi.org/10.1098/rstb.2009.0173

Page, M. P., Cumming, N., Norris, D., McNeil, A. M., & Hitch, G. J. (2013). Repetition-spacing and item-overlap effects in the Hebb repetition task. *Journal of Memory and Language*, *69*(4), 506-526. https://doi.org/10.1016/j.jml.2013.07.001

Page, M., Cumming, N., Norris, D., Hitch, G. J., & McNeil, A. M. (2006). Repetition learning in the immediate serial recall of visual and auditory materials. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(4), 716-733. https://psycnet.apa.org/doi/10.1037/0278-7393.32.4.716

Paice, A., Page, M. P. A., & Johnson, A. J. (2022). [Hebb effect and rhythmic stimuli: Learning underlying metrical patterns for lists of words]. Unpublished raw data.

Parmentier, F. B., Maybery, M. T., Huitson, M., & Jones, D. M. (2008). The perceptual determinants of repetition learning in auditory space. *Journal of Memory and Language*, *58*(4), 978-997. https://doi.org/10.1016/j.jml.2008.02.001

Poirier, M., & Saint-Aubin, J. (1995). Memory for related and unrelated words: Further evidence on the influence of semantic factors in immediate serial recall. *The*

Quarterly Journal of Experimental Psychology Section A, *48*(2), 384-404. https://doi.org/10.1080/002075999399675

Romani, C., Mcalpine, S., & Martin, R. C. (2008). Concreteness effects in different tasks: Implications for models of short-term memory. *Quarterly Journal of Experimental Psychology*, *61*(2), 292-323. https://doi.org/10.1080/17470210601147747

Saint-Aubin, J., & Guérard, K. (2018). The Hebb repetition effect as a laboratory analogue of language acquisition: Learning three lists at no cost. *Canadian Journal of Experimental Psychology*, *72*(1), 2-8. https://psycnet.apa.org/doi/10.1037/cep0000136

Saint-Aubin, J., & Poirier, M. (1999). Semantic similarity and immediate serial recall: Is there a detrimental effect on order information? *The Quarterly Journal of Experimental Psychology: Section A*, *52*(2), 367-394. https://doi.org/10.1080/713755814

Saint-Aubin, J., Ouellette, D., & Poirier, M. (2005). Semantic similarity and immediate serial recall: Is there an effect on all trials. *Psychonomic Bulletin & Review*, *12*(1), 171-177. https://doi.org/10.3758/BF03196364

Singmann, H., & Kellen, D. (2019). An introduction to mixed models for experimental psychology. In Spieler, D., & Schumacher, E. (Eds., pp. 5-28). (2019). *New methods in cognitive psychology*. Routledge.

Sjöblom, A., & Hughes, R. W. (2020). The articulatory determinants of verbal sequence learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *46*(10), 1977-1997. https://psycnet.apa.org/doi/10.1037/xlm0000926

Smalle, E. H., Bogaerts, L., Simonis, M., Duyck, W., Page, M., Edwards, M. G., & Szmalec, A. (2016). Can chunk size differences explain developmental changes in lexical learning? *Frontiers in Psychology*, *6*, 1925 https://doi.org/10.3389/fpsyg.2015.01925

Smalle, E. H., Szmalec, A., Bogaerts, L., Page, M. P., Narang, V., Misra, D., ... & Huettig, F. (2019). Literacy improves short-term serial recall of spoken verbal but not visuospatial items–Evidence from illiterate and literate adults. *Cognition*, *185*, 144-150. https://doi.org/10.1016/j.cognition.2019.01.012

St-Louis, M. È., Hughes, R. W., Saint-Aubin, J., & Tremblay, S. (2019). The resilience of verbal sequence learning: Evidence from the Hebb repetition effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 45*(1), 17-25. https://doi.org/10.1037/xlm0000566

Sukegawa, M., Ueda, Y., & Saito, S. (2019). The effects of Hebb repetition learning and temporal grouping in immediate serial recall of spatial location. *Memory & Cognition*, *47*(4), 643-657. https://doi.org/10.3758/s13421-019-00921-9

Szmalec, A., Duyck, W., Vandierendonck, A., Mata, A. B., & Page, M. P. (2009). The Hebb repetition effect as a laboratory analogue of novel word learning. *Quarterly Journal of Experimental Psychology*, *62*(3), 435-443. https://doi.org/10.1080%2F17470210802386375

Szmalec, A., Loncke, M., Page, M., & Duyck, W. (2011). Order or disorder? Impaired Hebb learning in dyslexia. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(5), 1270-1279. https://psycnet.apa.org/doi/10.1037/a0023820

Szmalec, A., Page, M. P., & Duyck, W. (2012). The development of long-term lexical representations through Hebb repetition learning. *Journal of Memory and Language*, *67*(3), 342-354. https://doi.org/10.1016/j.jml.2012.07.001

Yanaoka, K., Nakayama, M., Jarrold, C., & Saito, S. (2019). Determining the developmental requirements for hebb repetition learning in young children: Grouping, short-term memory, and their interaction. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*(4), 573-590. https://psycnet.apa.org/doi/10.1037/xlm0000606

Appendices

Appendix A1. Experiment 1 Stimuli Generation Study

Design

This study was designed to provide justification for the stimuli selected for the main study. The main study required six sets of 13 four-letter words, where each set comprised semantically related words. Here, 20 examples of each category were presented, with participants selecting the 12 best category exemplars for each set. The dependent variable is therefore the frequency of respondents who selected each word.

Participants

Fifty Bournemouth University Psychology students (aged 18-34, *M*=20.10, *SD*=3.10, 47 females) were recruited through the university's experiment participation scheme (Sona Systems, https://www.sona-systems.com) and compensated with course credits.

Materials

Stimuli consisted of 120 one-syllable, four-letter words, with 20 words belonging to each of the six semantic categories (Animals, Body Parts, Food, Music, Names, Places). The task was built and run using the experiment builder 'Testable' (https://www.testable.org/)

Procedure

After consent was obtained, participants completed six experimental trials. Each trial consisted of being shown 20 words, presented simultaneously in 4x5 grid. Each grid contained the 20 exemplars related to one category. Four versions of each grid were created, whereby the original grid was flipped vertically, then both grids were flipped horizontally, to ensure the same word was not consistently presented in the upper left corner. Participants were randomly presented with one of the four grids for each experimental trial. Participants then selected the 12 words that they thought were the best examples of each category. The task was self-paced, with participants typing their responses into the 12 empty text boxes positioned below the stimuli.

Results

Whilst participants were instructed to select the 12 best exemplars, a later iteration of the main experiment required 13 exemplar words, therefore, the analysis focussed on identifying the 13 most frequently selected exemplars for each category. It is worth noting that the word ‘Chad’ from the ‘Places’ category was removed from the final list due to potential overlap with another category (‘Names’). Therefore, ‘Looe’ was used at the thirteenth item from the ‘Places’ category (see Appendix A2 for frequency data).

Appendix A2. Frequency Table for Experiment 1 Stimuli Generation Study

Table 5.

Frequency of words chosen as the ‘best’ examples of each category. Note the gap between rows 13 and 14 show the cut-off point in which words rated in the top 13 were included as stimuli in Experiment 1. Words rated 14th and below were not included at stimuli except for the ‘Places’ category, in which Chad was removed and Looe was included based on Chad also falling into the category of ‘Names’.

Animals		Body Parts		Food		Music		Names		Places	
Word	Freq.	Word	Freq.	Word	Freq.	Word	Freq.	Word	Freq.	Word	Freq.
Goat	47	Foot	46	Cake	43	Jazz	48	Beth	45	Rome	49
Bear	46	Knee	45	Eggs	42	Drum	48	Jack	43	Kent	49
Wolf	46	Hand	45	Rice	41	Song	46	Kate	43	Hull	48
Deer	45	Head	42	Pork	40	Band	45	Jess	43	York	48
Bull	41	Neck	41	Milk	40	Rock	45	Luke	43	Bath	48
Seal	38	Back	41	Beef	40	Beat	41	Josh	42	Nice	43
Mole	35	Nose	40	Wine	36	Tune	40	Will	38	Hyde	42
Swan	33	Chin	37	Chip	35	Soul	40	Anne	36	Hove	40
Hare	33	Face	36	Soup	34	Folk	38	Mark	35	Cork	39
Frog	32	Lung	27	Beer	31	Harp	36	Pete	35	Chad	30
Toad	30	Limb	26	Pear	29	Hymn	36	Jane	32	Ryde	22
Hawk	28	Heel	26	Stew	28	Tone	31	Fred	29	Hook	21
Dove	24	Palm	24	Bean	25	Horn	19	Phil	21	Guam	20
Lynx	23	Shin	23	Corn	23	Clef	17	Zack	21	Looe	19
Crow	22	Nail	18	Brie	23	Bell	17	Troy	17	Holt	17
Mule	20	Skin	17	Plum	22	Pipe	13	Brad	17	Bern	14
Wasp	16	Hair	15	Kale	17	Gong	11	Fran	16	Lund	12
Slug	15	Bone	15	Lime	16	Rest	6	Elle	15	Omsk	12
Newt	8	Brow	9	Veal	12	Lute	6	Jill	13	Perm	8
Pike	4	Vein	7	Mint	3	Reed	6	Dale	5	Durg	7

Appendix B1. Experiment 2 Stimuli Generation Study

Design

This study was designed to provide justification for the stimuli selected for Experiment 2. This experiment required 36 category words, with two exemplars. For example, the category ‘Animal’ with ‘Dog’ and ‘Cat’ as exemplars. Here, three exemplars of each category were presented, with participants choosing the two they feel best match the category. The dependent variable is therefore the frequency of respondents who selected each word. The two words with the highest frequency from each category were then used as the stimuli for the main study.

Participants

Fifty Bournemouth University Psychology students (aged 18-50, *M* = 21.34, *SD* = 6.58, 37 females) were recruited through the university’s sona system (https://www.sona-systems.com) and compensated with course credits. None had participated in Experiment 1 or the preceding pilot study.

Materials

Stimuli consisted of 36 category words (such as ‘season’, ‘herb’, ‘snake’) and 108 exemplars (such as ‘Autumn’, ‘Basil’, ‘Python’). Category and exemplar words were chosen based on being between 3-9 letters, and either one or two syllables. For every category word, there were three exemplars. The task was built and run using the experiment builder ‘Testable’ (www.testable.org).

Procedure

After consent was obtained, participants completed 36 experimental trials. Each trial consisted of being presented with one category word, with the three exemplars below and two empty text boxes. Participants typed the two words they felt were the best examples of the category into the text boxes. On completion of each trial, participants pressed the ‘return’ key. The task was self-paced, and the order of the categories and the position of each exemplar was randomised.

Results

The frequency with which words were selected were calculated and the two words with the highest frequencies for each category were used as stimuli for the main study. Frequencies are reported in Appendix B2.

Appendix B2. Frequency Table for Experiment 2 Stimuli Generation Study

Table 6.

Frequency of words chosen as the 'best' examples of each category. Note that the words in the first columns have the highest frequency scores for each category and words in the third column had the lowest frequency scores. Stimuli for Experiment 2 were taken from the first two columns (i.e., the two most frequently chosen exemplars).

Category	Exemplars					
	Word	Freq.	Word	Freq.	Word	Freq.
Flower	Tulip	42	Daisy	30	Orchid	27
Season	Winter	43	Summer	30	Autumn	26
Planet	Saturn	34	Venus	34	Neptune	30
Fruit	Apple	47	Orange	26	Pear	24
Colour	Red	37	Yellow	36	Green	24
Drink	Tea	40	Coffee	40	Water	18
Music	Reggae	34	Rap	33	Jazz	32
Dance	Salsa	48	Ballet	28	Rumba	22
Insect	Beetle	44	Earwig	42	Aphid	13
Sport	Football	43	Rugby	41	Tennis	13
Clothing	Jumper	39	trousers	33	Dress	25
Dairy	Cheese	37	Milk	32	Butter	26
Shape	Square	47	Circle	38	Rhombus	12
Snake	Cobra	42	Python	38	Boa	17
Tree	Oak	49	Pine	37	Elm	13
Curry	Korma	37	Madras	37	Bhuna	25
Herb	Basil	42	Parsley	32	Sage	25
Fish	Salmon	44	Tuna	39	Haddock	15
Bird	Pigeon	39	Seagull	37	Eagle	21
Boat	Yacht	45	Ferry	42	Canoe	12
City	Paris	46	London	37	Madrid	15
Film	horror	46	Sci-Fi	39	Romcom	14
Rodent	Rat	47	Mouse	43	Squirrel	7
Medic	Doctor	49	Nurse	29	Surgeon	20
Hair	Blonde	44	Brunette	42	Ginger	13
Language	French	41	Spanish	34	English	24
Hat	Beret	47	Turban	31	Trilby	21
Game	Scrabble	45	Charades	28	Poker	24
Bread	Baguette	41	Bagel	35	Sourdough	19
Meat	Pork	36	Ham	30	Beef	29
Number	Nine	42	Seven	41	Four	16

Category	Exemplars					
	Word	Freq.	Word	Freq.	Word	Freq.
Career	Lawyer	46	Banker	26	Teacher	25
Tool	Hammer	49	Drill	36	Pliers	14
Metal	Copper	41	Steel	33	Brass	25
Nut	Almond	43	Cashew	38	Pecan	16
Month	July	43	June	42	April	12