University of Kent

Kent Academic Repository

Ishikawa, Misae (2021) ***The Effects of Task-Irrelevant Information on Memory: The Associative Memory Stroop Task.*** **Doctor of Philosophy (PhD) thesis, University of Kent,.**

Downloaded from

https://kar.kent.ac.uk/86560/ The University of Kent's Academic Repository KAR

The version of record is available from

https://doi.org/10.22024/UniKent/01.02.86560

This document version

UNSPECIFIED

DOI for this version

Licence for this version

CC BY (Attribution)

Additional information

Versions of research works

Versions of Record

If this version is the version of record, it is the same as the published version available on the publisher's web site. Cite as the published version.

Author Accepted Manuscripts

If this document is identified as the Author Accepted Manuscript it is the version after peer review but before type setting, copy editing or publisher branding. Cite as Surname, Initial. (Year) 'Title of article'. To be published in *Title of Journal* , Volume and issue numbers [peer-reviewed accepted version]. Available at: DOI or URL (Accessed: date).

Enquiries

If you have questions about this document contact ResearchSupport@kent.ac.uk. Please include the URL of the record in KAR. If you believe that your, or a third party's rights have been compromised through this document please see our Take Down policy (available from https://www.kent.ac.uk/guides/kar-the-kent-academic-repository#policies).

The Effects of Task-Irrelevant Information on Memory:

The Associative Memory Stroop Task

Misae Ishikawa

School of Psychology

University of Kent

2020

72,282 words, 257 pages

In part fulfilment of the requirements of the Degree of Doctor of Philosophy

Acknowledgements

I would first like to express the deepest appreciation to my supervisor, Dr Dinkar Sharma. His valuable guidance taught me how I needed to choose the right direction and successfully complete my research. Without his patience, advice and encouragement, this thesis would not have been possible.

I would also like to sincerely thank the technical support team of School of Psychology at University of Kent for their help and providing me with all the necessary facilities. Their immense knowledge has made me become interested in computer language, which I thought I could never have an interest in before.

A heartfelt thank you to the undergraduate students who helped my research through the Research Experience Scheme (RES). Conducting research with them helped me to develop not only my research abilities but also leadership abilities.

I would also like to say a very special thank you to my PhD colleagues whose warm-hearted words and kindness were emotional support for me while staying in a foreign country. Thank you for always being wonderful friends.

Finally, a big thank you to my family for their continued inspiration support. Without their understanding and encouragement, I would not have had this precious experience of studying abroad.

Abstract

The goal of the present thesis was to evaluate the validity of a new Stroop-like paradigm as a tool to measure paired-associate learning. Hazan-Liran and Miller (2017) developed a task to study how task-irrelevant ink colour information affects learning efficacy (see also Miller, Hazan-Liran, & Cohen, 2018). The task asks participants to learn pairs of colour concepts and numbers (e.g. *blue-5*) whilst the concepts are presented in a neutral ink colour (e.g. black ink) but the numbers are coloured in either congruent ink colours (e.g. the number *5* printed in blue ink), incongruent ink colours (e.g. *5* in brown ink) or neutral ink colours (e.g. *5* in black ink). The present thesis named this task the Associative Memory Stroop Task (AMST) and aimed to replicate the findings in Hazan-Liran and colleagues' studies when task-irrelevant ink colour effects on this task are measured by memory accuracy on cued-recall tests and recognition tests. Colour names (e.g. *blue*), colour-related words (e.g. *sky*) and emotional words (e.g. *unhappy*) were employed. It was hypothesised that more accurate memory performance would occur with congruent ink colours than neutral ink colours (i.e. facilitation) whilst poorer memory accuracy would be reported with incongruent ones compared to neutral ink colours (i.e. interference). Chapter 2 tested colour names and colour-related words with cued-recall tests and recognition tests. Chapter 3 explored the use of colour-related concepts on the AMST with recognition tests. Chapter 4 examined the emotional factor during the AMST with cued-recall tests. Overall, the analyses revealed facilitation and interference effects on cued-recall tests and interference on recognition tests when the words were colour names. Other types of colour words failed to produce the effects. The present research suggested possible explanations for these results.

Contents

Chapter 1: Theoretical Frameworks and Issues

This chapter aims to summarise reported evidence of paired-associate learning, the Stroop effect, cognitive load theory and the two studies that investigated the Stroop-like effects in paired-associate learning with considering cognitive load. The goal of this chapter is to demonstrate the rationale for why it is possible to speculate that the findings of the studies conducted by Hazan-Liran and colleagues (i.e. task-irrelevant ink colours affect learning efficacy; Hazan-Liran & Miller, 2017; P. Miller et al., 2018) can be applied to memory performance. To achieve this goal, Chapter 1 first introduces a concept of paired-associate learning, which occurs frequently in our daily lives, and how this type of learning can be enhanced (i.e. **Associating Two Concepts as a Pair**). Then the unique phenomenon that occurs due to the memories created by paired-associate learning is demonstrated (i.e. **The Effects of Pre-existing Colour-Word Associations**) and the potential contributors of such a phenomenon are clarified (i.e. **Conflict in the Stroop Effect**). The present research approaches the underlying processes of the phenomenon from a mental load perspective (i.e. **Cognitive Load Theory**). As this may raise a question for whether previously-learned colour-word associations affect paired-associate learning, two studies that investigated this issue are introduced (i.e. **Hazan-Liran and Colleagues' Studies**). By discussing the importance and limitations of their research (i.e. **Limitations of the Two Original Studies**), the research question, aim and hypotheses of the current thesis are presented (i.e. **Overview of the Present Research**).

Associating Two Concepts as a Pair

When you are trying to remember a new person's face and name, this is called paired-associate learning. Specifically, this is the process of newly creating a link between two concepts (for an overview see Arndt, 2012). This type of learning occurs frequently not only at school but also in our daily lives such as when trying to remember how much one's favourite brand's new product is from the advertisement. Therefore, investigation of paired-associate learning would deliver a benefit to both academia and outside academia.

There are several models to explain why such associations can occur. For example, it might occur since stimuli that appear closely each other tend to be linked together or since the items that are stored with the same contextual information in memories are associated (for an overview see Davis, Geller, Rizzuto, & Kahana, 2008). Through paired-associate learning, a memory for the connected concepts is created. This is named an associative memory (Matzen, Trumbo, Leach, & Leshikar, 2015). One might question whether paired-associate learning can occur without forming this memory. The studies with brain studies would provide an answer for this question; paired-associate learning seems to produce a change in memory. During paired-associate learning, the involvement of certain brain areas has been reported such as frontal lobe (e.g. Shimamura, Jurica, Mangels, Gershberg, & Knight, 1995) and hippocampus (e.g. Rajji, Chapman, Eichenbaum, & Greene, 2006). The literature reported that these brain areas relate to creation of associative memories (Hales & Brewer, 2011; Nakazawa et al., 2002). Since associative memories are the products from paired-associate learning, the quality of paired-associate learning tends to be examined with memory tests such as cued-recall tests or associative recognition tests (Arndt, 2012).

As paired-associate learning occurs in our daily lives frequently, it is reasonable to question how the quality of this learning is modified. One factor that affects such a quality is word concreteness. For example, increasing word concreteness was reported to improve accuracy of paired-associate learning (e.g. Paivio, 1965). However, when it is impossible to change the intrinsic features of learning materials, sleep can help paired-associate learning (e.g. Fenn & Hambrick, 2012, 2013; Potkin & Bunney, 2012). The literature (for a summary see Campos, Barroso, & de Lara Menezes, 2010) found that through sleep, short-lived memories can be transformed into long-lived memories. This transformation of memories is called consolidation (Diekelmann & Born, 2010; Klinzing, Niethard, & Born, 2019). Whilst it is still controversial the underlying mechanisms for enhancement of paired-associate learning through sleep, it may occur due to protection of memories from loss that usually occurs during wakefulness (Fenn & Hambrick, 2013). The literature demonstrated the benefit of sleep consolidation on memory durability; therefore, one might question how it delivers a benefit on paired-associate learning. If you forget what you remember immediately, you would have to learn it every time when you need such information. By transforming into more durable memory, it helps to not only obtain information but also make newly learned information accessible later. Thus, sleep can deliver a benefit to paired-associate learning. Also, changing the stimulus presentation format (e.g. presenting as a picture instead of words) has been reported to improve paired-associate learning (Paivio & Yarmey, 1966). Whether each member of a pair induces any emotional reactions from learners also modify performance of paired-associate learning. For example, when one member of a pair triggers negative emotions, learners can remember this item accurately whilst their memory for the paired item is impaired (Bisby & Burgess, 2014; Bisby, Horner, Hørlyck, & Burgess, 2016).

To summarise, whilst this type of learning occurs frequently in daily lives, performance of it is sensitive to intrinsic features of learning materials (i.e. concept concreteness, presentation formats and affective valences) and learners' circumstance (i.e. experience of sleep). Since the aforementioned research demonstrated this type of learning can be improved intentionally (e.g. presenting materials as pictures), this would suggest that the goal of paired-associate learning research is finding a way to optimise this type of learning. As these studies have demonstrated evidence for when forming a new association between two items, this may raise a question; how established associations affect subsequent behaviour. The pre-existing association of semantically unrelated word-word pairs accelerated the subsequent word naming when these words were consecutively presented as a pair (Spieler & Balota, 1996). Whilst this demonstrated the effects of pre-existing associations, one can expect that creation of a new link would be more likely to occur between semantically related concepts such as between blue ink and the word *blue*. Such previously-established links between ink colours and colour words have produced an interesting phenomenon when naming the ink colours of colour words.

The Effects of Pre-existing Colour-Word Associations

Stroop's (1935) land mark study that developed the new paradigm called the Stroop task has been used to study attention (for a review see MacLeod, 1991). In the traditional Stroop task there are three colour congruency conditions: congruent, incongruent and neutral. In the congruent condition ink colours of the colour words match the meanings (i.e. semantics) of the words (e.g. the word *RED* printed in red ink) whilst in the incongruent condition the ink colours did not match (e.g. the word *RED* printed in green ink). The stimuli can be considered neutral when these do not activate specific colour-word associations such as the colour words printed in a neutral ink colour (e.g. black ink) or the randomly-coloured squares. The task tends to be colour naming of the presented colour words whilst ignoring the meanings of the words. By comparing task performance (usually on a reaction time dimension) between the three colour congruency conditions, the Stroop task allows researchers to examine how colour-word congruency affect cognitive processes. When any differences in task performance are observed, this phenomenon is called *the Stroop effect*; readers process the meanings of colour words on a colour-naming task although processing such information is irrelevant to the task (De Marchis, Rivero Expósito, & Reales Avilés, 2013). The Stroop effect can be divided into two types depending on between in which conditions task performance differs. When participants' response is accelerated with congruent ink colours compared to the neutral condition, this is considered as *the Stroop facilitation* whilst *the Stroop interference effect* occurs when their reaction time becomes slower with incongruent ink colours than with the neutral ink colours (De Marchis et al., 2013). These effects are usually interpreted using attentional models (Besner & Stolz, 1999; MacLeod, 1991). Thus, the Stroop effect can be either advantageous (i.e. facilitation) or harmful (i.e. interference). Additionally, the robustness of the Stroop effect is affected by how strongly colours and words are related. For example, the word *RED* activates the colour image of red more strongly than the word *TOMATO* since the former word triggers the colour image directly (Levin & Tzelgov, 2016). Since the activation of the concept becomes weaker as it passes through more concepts (the semantic network model; Collins & Loftus, 1975), the word *TOMATO* tends to produce the weaker Stroop effect than the word *RED*.

Conflict in the Stroop Effect

The traditional view for the source of the Stroop effect has been that conflict (i.e. competition that occurs mentally) contributes to this effect. According to the two-conflict framework (Goldfarb & Henik, 2007; Levin & Tzelgov, 2016; MacLeod & MacDonald, 2000), task conflict and/or informational conflict trigger(s) the Stroop interference effect. Task conflict occurs when two different tasks (i.e. word reading and colour naming) compete each other (Goldfarb & Henik, 2007; Henik, Bugg, & Goldfarb, 2018; Levin & Tzelgov, 2016; MacLeod & MacDonald, 2000). Because it is more natural to read words rather than to name the colours of the ink in which words are written, people have to control the automatic tendency to read words in the Stroop task (Entel, Tzelgov, Bereby-Meyer, & Shahar, 2015). Therefore, it becomes slow to name the ink colours of the colour words with incongruent ink colours. However, as Entel et al. noted, this concept may raise a question for why performance with congruent and incongruent ink colours differs since naming both ink colours triggers competition between the automatic word-reading behaviour and colour naming. To clarify this issue, another type of conflict was introduced; informational conflict. Informational conflict refers to a disagreement between the activated information from ink colours and from words (Entel et al., 2015; MacLeod, 1991). For example, when the word *GREEN* is written in blue ink, the information that *GREEN* activates is different from the information that blue ink triggers. Shichel and Tzelgov (2018) proposed that informational conflict can be divided into two types: response conflict and semantic conflict. Response conflict refers to the assignment of different responses to each colour. For example, participants need to press the enter key for red ink and the space key for blue ink. Since the required response differs depending on each colour, it creates confusion and leads to a slower reaction. In this way, response conflict is different from task conflict (i.e. a competition between a colour-naming task and a word-reading action) due to the relevance to ink colour congruency. Shichel and Tzelgov (2018) also suggested another type of informational conflict; semantic conflict. According to them, this conflict occurs when activated "meanings" are different. Shichel and Tzelgov exemplified this concept with the hypothetical example of the word *RED* in blue ink. In this case, reading the word *RED* induces the colour memory of "red." However, the task is to name the ink colours. Thus, participants need to analyse the ink colour information, which is blue in this case. Because the activated colour memory (i.e. "red") is different from the ink colour (i.e. blue), this colour memory hampers the analysis of the ink colour (see also Hock & Petrasek, 1973); Consequently, the incongruent ink colours produce slower response speed.

Whilst previously-established colour-word associations produce affects a colour naming speed, it would be reasonable to question how these affect other cognitive tasks. If the effects are limited to colour-naming behaviour, there would be no reason to consider such associations important when engaging in different tasks (e.g. word reading). Based on the suggested relationship between written words and word reading (e.g. Entel, Tzelgov, Bereby-Meyer, & Shahar, 2015), this can indicate that colour naming occurs less frequently than word reading in our daily lives. Therefore, one could question about the applicability of the Stroop effect in a real-life setting. However, whilst the Stroop literature (MacLeod, 1991; Stroop, 1935) demonstrated that pre-existing colour-word associations had no impact on a word reading speed, such links do have an effect on readers' emotions even when they simply see the words.

Interestingly, the occurrence of conflict due to colour-word incongruency seems to trigger emotional reactions. For instance, it was found that simply seeing the incongruent Stroop stimuli (e.g. the word *RED* printed in blue ink) makes people feel displeasure (Damen, Strick, Taris, & Aarts, 2018) because people see conflict as negative events (Botvinick, 2007; Dreisbach & Fischer, 2012). The research of the emotional influence triggered by the Stroop stimuli has allowed researchers to investigate such an effect from the new dimension. Since only colour and word dimensions have been the main focus of the traditional Stroop research, this recent research trend of how the Stroop stimuli affect one's emotions has successfully provided support for the applicability of the Stroop effect on a new dimension.

To summarise, it was illustrated that pre-learned colour-word associations can produce meaning conflict (i.e. semantic conflict) between ink colours and colour words and this leads to interference on a colour-naming task. Although the literature explained that conflict contributes to the Stroop interference effect, one might question "why" conflict produces such an effect. Whilst the traditional Stroop literature has mainly focused on the root of the Stroop effect, Hazan-Liran and colleagues' (Hazan-Liran & Miller, 2017; P. Miller et al., 2018) cognitive load theory approach to investigate such an effect seems to be interesting. That is, whilst the amount of available cognitive resources is limited, dealing with conflict in one's head uses up such limited resources and results in less available resources to complete a task. As the present thesis attempts to explain the Stroop-like effects based on their view point, the following section introduces the main theory: cognitive load theory.

Cognitive Load Theory

The studies conducted by Hazan-Liran and colleagues (Hazan-Liran & Miller, 2017; P. Miller et al., 2018) were framed around the cognitive load theory, which proposes the importance of cognitive load to optimise the quality of learning (Sweller, 1988, 1994; Sweller, Ayres, & Kalyuga, 2011; Sweller & Chandler, 1994). The pivotal assertion of cognitive load theory is that since the capacity and duration of working memory are limited (Cowan, 2001; Hazan-Liran & Miller, 2017; G. A. Miller, 1956; Peterson & Peterson, 1959; Sweller et al., 2011), the overload of working memory capacity leads to the interference effect on learning (de Jong, 2010). However, when acquired information can be combined via schemas (i.e. the mental framework that organise information; van Merriënboer & Ayres, 2005), this saves the amount of elements that is retained in working memory. Consequently, it allows to learn much information without the overload (Debue & van de Leemput, 2014; Sweller et al., 2011). According to the cognitive load theory, whether information can be combined together as one schema or not contributes to the degree of *intrinsic cognitive load*. Specifically, cognitive load theory assumes that the level of intrinsic cognitive load involves learners' level of expertise since more expertise learners can combine more elements into one element (Debue & van de Leemput, 2014; Hazan-Liran & Miller, 2017; Kalyuga, 2007; P. Miller et al., 2018; Schnotz & Kürschner, 2007; Sweller, 2010; Sweller et al., 2011; Sweller, van Merriënboer, & Paas, 1998; Sweller & Chandler, 1994; van Merriënboer & Ayres, 2005; van Merriënboer & Sweller, 2005). Whilst the amount of this type of cognitive load depends on learners' abilities, an external factor also determines another type of cognitive load: *extraneous cognitive load*.

Extraneous cognitive load is related to the processes of task-irrelevant information (Debue & van de Leemput, 2014; Hazan-Liran & Miller, 2017; Kalyuga, 2007; P. Miller et al., 2018; Schnotz & Kürschner, 2007; Sweller, 1994, 2010; Sweller et al., 2011, 1998; Sweller & Chandler, 1994; van Merriënboer & Ayres, 2005; van Merriënboer & Sweller, 2005). This type of cognitive load is the one that the AMST attempts to control. The amount of extraneous cognitive load is determined by the stimulus presentation format or engagement in irrelevant cognitive activities (for a review see Schnotz & Kürschner, 2007). Note that the amount of intrinsic cognitive load and of extraneous cognitive load are additive (Schnotz & Kürschner, 2007). Therefore, when the amalgam of intrinsic and extraneous cognitive loads exceeds the available amount of working memory resources, the cognitive system becomes unable to function that can indicate students' failure of learning (Sweller et al., 2011). Whilst the amount of intrinsic cognitive load is determined by learners' expertise, only the level of extraneous cognitive load is flexible from teachers' or instructors' perspective. Therefore, minimising extraneous cognitive load is the key to improve learning efficacy (Sweller et al., 2011). In the case of the stimulus presentation format, the inappropriate format can increase the level of extraneous cognitive load and interferes with learning. Sweller et al. (2011) described how such an external factor may modify students' learning performance by using the figures of a geometry problem (see Figure 1.1).

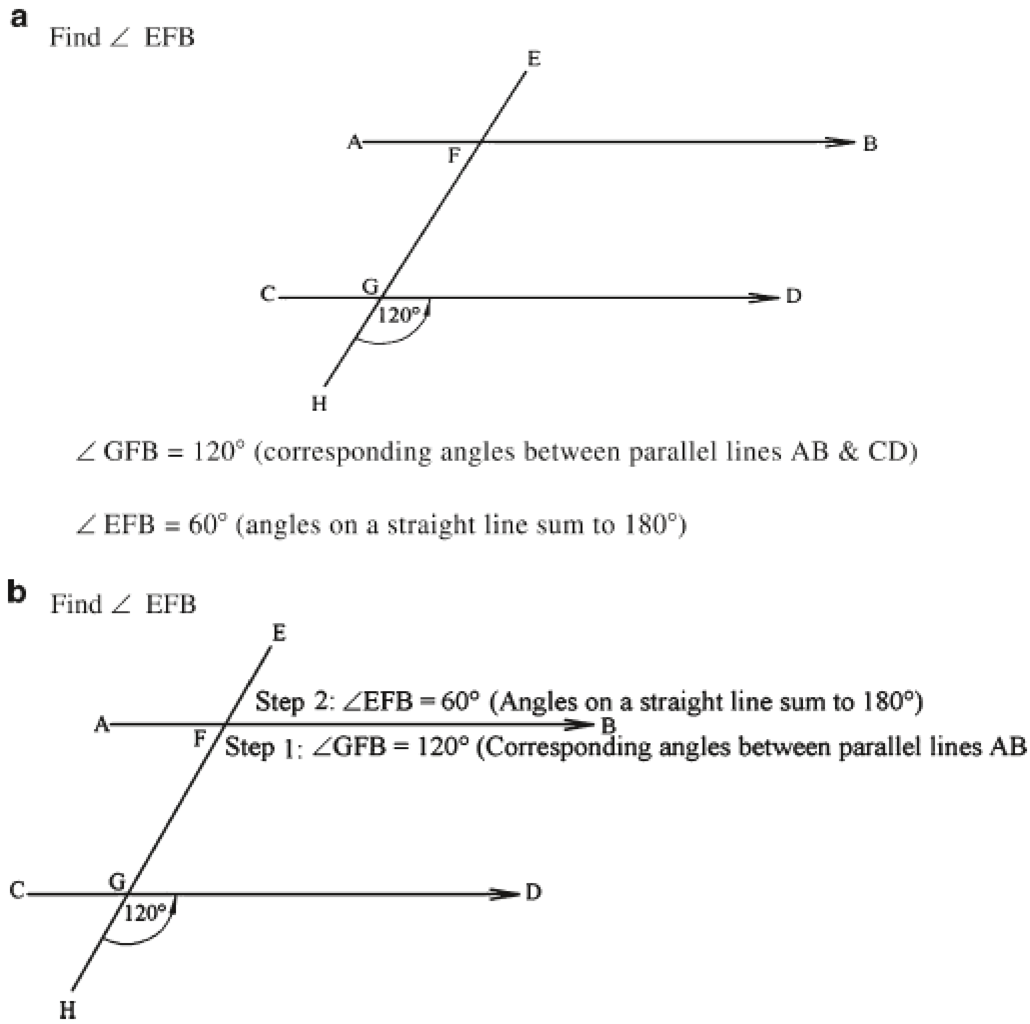


Figure 1.1. These figures are retrieved from Sweller et al. (2011) and represent the same geometry problem (which asks to calculate the angle of EFB) with the two-step solution. (a) This picture produces high extraneous cognitive load since the pictorial presentation of the problem and the written solution are presented separately, which requires learners to shift their attention between the picture and the written solution. (b) This picture combined the information on the picture with the one in the written solution in the easier way, which results in the reduction of extraneous cognitive load.

Figure 1.1 represents a simple geometry problem. The task is to calculate the angle of EFB and the written two-step solution is provided. However, the location of this solution contributes to the level of extraneous cognitive load. As Sweller et al. (2011) emphasised, learning is optimised when learners can mentally combine all pieces of information. As an aid to help them to combine information, Figure 1.1a shows an inappropriate information presentation whilst Figure 1.1b seems to be more appropriate. In Figure 1.1a the pictorial presentation of the problem is presented at the top and the two-step solution is presented at the bottom. Since the solution to answer the problem is presented separately from the picture, learners need to shift their attention between the picture and the written solution at the bottom. This attention shift hampers them to combine the information in the picture with the one in the written solution and is irrelevant to the task (i.e. calculate the angle of EFB). Consequently, the level of extraneous cognitive load is increased. However, this increment of extraneous cognitive load can be prevented by modifying the way of presenting information. Figure 1.1b exemplifies the more efficient way to present materials, which helps students to combine the information in the picture and in the written solution by presenting them closely each other.

Another circumstance when extraneous cognitive load occurs is a case of engaging in irrelevant cognitive activities. When learners use cognitive resources to perform mental activities that is irrelevant for learning goals, this use of the resources are also considered as increasing extraneous cognitive load (Kalyuga, 2007; also see Schnotz & Kürschner, 2007). Note that the inappropriate stimulus presentation format also seems to trigger this pattern of extraneous cognitive load (Kalyuga, 2007). The present research suggests that extraneous cognitive load that is produced on the AMST fits this pattern since the main source of cognitive load on the AMST would be suppression of task-irrelevant information as Hazan-Liran and Miller (2017) described. That is, when suppression of task-irrelevant information diverts cognitive resources from learning the word-number pairs, learning efficacy is impaired although this inhibition is an additional process for a task goal (i.e. learning the pairs). The AMST triggers this additional process by presenting the learning materials with incongruent ink colours; the stimulus presentation format manipulates the required amount of cognitive resources for learning. Consequently, the present thesis assumes that controlling ink colour congruency would activate engagement of task-irrelevant cognitive activities, which produce extraneous cognitive load.

Despite the reports of harmful effects from high extraneous cognitive load on learning (e.g. Hazan-Liran & Miller, 2017; P. Miller et al., 2018), one might question why in some circumstances, inserting supplemental figures or graphs help students' comprehension of learning materials although the process of these graphical aids is redundant. This question can be solved by considering whether the function of information and the engagement in cognitive activities fit learners' level of expertise (de Jong, 2010). According to Sweller et al. (2011), supplemental materials help beginners' learning whilst processing these may generate the negative effects on advanced learners (i.e. those who have already obtained information or knowledge). When advanced learners are required to process the extra materials (which delivers the same information as the main learning materials), this is the situation that leaners process the same information repeatedly although they have already acquired the main information. Due to this redundant process of information that is irrelevant for learning, extraneous cognitive load is imposed. Consequently, inserting figures or graphs consumes advanced learners' cognitive resources and produces the detrimental effect. This is not the case for beginners. Since they have not had essential information or knowledge yet, processing the same information can be advantageous when this process aims to help learning. For example, Schnotz and Rasch's (2005) study demonstrated that adding the static pictures enhanced low-experience learners' comprehension more effectively than the animated pictures whilst high-experience students' learning efficacy was identical regardless of the type of the supplemental materials. Sweller et al. (2011) explained that since animation is more complex and cognitively demanding than static pictures, low-experience learners were unable to manage information in animations and thus learned better with static pictures. Therefore, processing additional information can be beneficial or harmful depending on learners' level of skills and knowledge.

The application of the main assumption of cognitive load theory to the Stroop effect would be important as it may provide an additional explanation for why previously-established colour-word associations produce interference. The main point of the theory is that processing task-irrelevant information or engaging task-irrelevant cognitive activities uses up the limited cognitive resources. Thus, the present thesis assumes the following scenario; on a traditional Stroop task, word meanings are automatically processed although such information is irrelevant to the task. Due to its irrelevancy to the task, participants need to suppress this information that requires additional cognitive resources (Hazan-Liran & Miller, 2017); thus, the less resources are left for completing the main colour-naming task. Whilst the traditional Stroop literature has demonstrated that conflict contributes to the Stroop effect, the approach from cognitive load theory view seems to clarify what happens in one’s head when colour-word incongruency occurs. Additionally, as cognitive load theory has its root in the field of learning, this could signal the possible approach for investigating the effects of pre-existing colour-word associations in the area of learning.

When trying to learn materials, it is common to use colours. For example, highlighting the important sentences in the textbook or presenting these in different ink colours (e.g. red ink) would be the common techniques to learn materials more effectively. Although one can intuitively see how frequently colours are used during learning, evidence of how previously-learned colour-word associations affect learning efficacy was insufficient until Hazan-Liran and colleagues (Hazan-Liran & Miller, 2017; P. Miller et al., 2018) conducted research. Traditionally, when colour-word associations were examined, participants needed to process ink colour information to name the ink colours. As processing such information activates the associated colour names (Hazan-Liran & Miller, 2017), this would indicate that colour-word associations are relevant to the colour-naming task. However, during learning, word reading would be required more frequently than colour naming. Since pre-existing colour-word associations have shown no impact on a word reading speed, this may raise a question; when learners do not need to process ink colours during learning, whether previously-established colour-word relationships affect learning efficacy. An answer for this question may provide benefits in two ways. First, it would demonstrate the way to optimise learning performance (i.e. the goal of learning research). Second, whilst the traditional Stroop research focuses on the effects of task-relevant colour-word links, such an answer would show whether task-irrelevant colour-word associations modifies human behaviour. Therefore, the investigation of how contextual colour information and triggered colour-word associations during learning would provide additional evidence not only to the field of learning but also to the area of the Stroop research. Since Hazan-Liran and colleagues' (Hazan-Liran & Miller, 2017; P. Miller et al., 2018) research investigated this important issue, the present research aims to extend their works by overcoming their limitations.

Hazan-Liran and Colleagues' Studies

Importantly, Hazan-Liran and colleagues' (Hazan-Liran & Miller, 2017; P. Miller et al., 2018) studies demonstrated that task-irrelevant colour-word associations modified performance of paired-associate learning. Their research aimed to examine when a pair of colour words and numbers is a to-be-learned material, whether colouring the numbers with congruent ink colours or incongruent ink colours (e.g. blue-7 or blue-7) changes the easiness of learning. The development of this new paradigm was important since by setting ink colours and its relationships to colour words as task irrelevant, this paradigm allows researchers to investigate the effects of extraneous cognitive load imposed by pre-existing colour-word associations. The present thesis named their paradigm to examine such effects as the Associative Memory Stroop Task (AMST). The effects of processing such irrelevant information on paired-associate learning were measured by the number-copy task, which asked to copy the digits from the example list on top of the page (see Figure 1.2).

a

brown	orange	white	black	green	red	yellow	blue
6	2	5	1	3	8	4	7

White	red	blue	orange	red	white	orange	blue	brown	yellow	black	green
5	8	7	2	*8	*5	*2					

b

brown	orange	white	black	green	red	yellow	blue
7	1	8	6	4	2	5	3

white	red	blue	orange	red	white	orange	blue	brown	yellow	black	green
8	2	3	1	*2	*8	*1					

c

brown	orange	white	black	green	red	yellow	blue
5	3	2	6	4	7	1	8

white	red	blue	orange	red	white	orange	blue	brown	yellow	black	green
2	7	8	3	*7	*2	*3					

Figure 1.2. The examples of the materials retrieved from Hazan-Liran and Miller (2017). Figure 1.2a represented the congruent condition, Figure 1.2b demonstrated the incongruent condition and Figure 1.2c was the neutral condition. The task was the number-copy task that asked the participants to check the example word-number pair in the example list (i.e. the lists coloured in grey) and to write down the correct paired digit in the blank response box below. In the bottom rows of the response section, the first four response boxes were completed by the experimenter as a demonstration. The asterisk represented the participants' actual responses.

By analysing the number of copied digits within 2 minutes, the highest number of completion was observed with the congruent ink colours whilst the lowest occurred with the incongruent ink colours. They observed this result with colour names and colour-related words (e.g. *sky*, *strawberry*, etc.) and confirmed that colour-word incongruency on the AMST contributed to interference on paired-associate learning (P. Miller et al., 2018). The uniqueness of their research would be that explaining such effects of task-irrelevant colour-word associations by introducing cognitive load. That is, seeing ink colours triggered the associated colour names. When this activated colour names mismatched the to-be-learned colour words, learners needed to suppress this redundant colour name information. Since this inhibition used up the limited cognitive resources, the less amounts of resources were left for forming a new link between colour words and digits. Consequently, the quality of paired-associate learning with the AMST was degraded. Since neutral ink colours did not require such inhibition, performance of paired-associate learning was not affected. When congruent ink colours were presented, the activated colour names matched the to-be-learned colour words. By repeated activation of the same word information, it helped learners to remember the learning materials.

The importance of their research would be highlighted by considering their significant contributions in three areas: paired-associate learning, the Stroop research and cognitive load theory. In the field of paired-associate learning research, the goal of this area would be finding a way to optimise this type of learning. Since how ink colours affect learning performance was unclear despite their frequent use in learning, Hazan-Liran and colleagues' research provided significant evidence that how such common elements can optimise learning efficacy. Second, as the traditional Stroop task has focused on the effects of task-relevant colour-word links on a colour naming speed, their development of the new paradigm extended the applicability of the traditional Stroop-like effects. Specifically, they successfully provided evidence that the effects of pre-existing colour-word associations are not limited to a colour naming behaviour. Their research results demonstrated that simply seeing the Stroop-related stimuli can modify one's cognitive performance. Finally, since how to measure the effects of cognitive load has been a controversial issue in the area of cognitive load theory (for a summary see de Jong, 2010), evidence with the new paradigm that measures the impacts of cognitive load would help to solve this issue.

As these illustrated, Hazan-Liran and colleagues' (Hazan-Liran & Miller, 2017; P. Miller et al., 2018) research provided important evidence. However, establishing the credibility of these contributions would need the confirmation of the ability of the AMST to affect paired-associate learning. Despite the importance of their research, the limitations in their methodology made their findings with the AMST questionable. Specifically, their use of the number-copy task to measure the effects of colour-word associations in the AMST seemed to raise some questions about whether their findings were truly related to learning efficacy. If their findings occurred due to other processes (e.g. visual search strategy), it would be difficult to conclude that task-irrelevant colour-word associations produced in the AMST affect the quality of paired-associate learning. Therefore, it would be important to discuss their limitations and how the present thesis attempts to overcome these.

Limitations of the Two Original Studies
(Hazan-Liran & Miller, 2017; P. Miller et al., 2018)

The findings of the two original studies (Hazan-Liran & Miller, 2017; P. Miller et al., 2018) provided beneficial implications of how manipulation of task-irrelevant semantic relationships between contextual dimension and learning materials affects learning efficacy. However, despite their significant contribution, the findings are limited to the validity of their measurement method of learning; the aspects of learning, the authenticity of learning and the root of the observed effects on learning.

For Hazan-Liran and Miller's (2017) and P. Miller et al.'s (2018) conclusions, the present thesis suggests the alternative interpretation of their findings; that is, although the researchers (Hazan-Liran & Miller, 2017; P. Miller et al., 2018) generalised their findings to the overall aspects of learning efficacy, it is possible to consider their results as the consequence of learning speed differences. This possibility arises from their analyses of the dependent variable; the number of completed response boxes within 2 minutes. Under the pressure of the time limits, the participants were asked to fill out the response boxes as fast and accurately as possible. Therefore, this instruction seemed to encourage them to develop the strategy in order to accelerate their speed of completing the test sheet. For the number-copy task, the researchers' (Hazan-Liran & Miller, 2017; P. Miller et al., 2018) basic assumption was that as the participants learned the word-number pairs more efficiently, they did not need to check the example list and could fill out the response boxes directly based on their working memory (D'Eredita & Hoyer, 2010). Due to this elimination of searching processes in the example list, the participants could use this time to complete more response boxes; thus, the number of completions reflected learning efficacy. However, it is noteworthy to highlight that there was no constraint on information accuracy of word-number pairs at the stage of inputting information. Since the example list was available on top of the page, the participants could check the example word-number pairs at any time during the task, which indicates a lower demand on remembering word-number pairs accurately on the first try. Therefore, although the researchers (Hazan-Liran & Miller, 2017; P. Miller et al., 2018) did not specify on what aspect of learning the manipulation of task-irrelevant ink colours in the AMST affects, it is possible to presume that their findings of ink colour effects might have the impact on learning speed dimension rather than learning accuracy aspect. Due to this, evidence from their studies seems to be insufficient to determine whether the ink colour manipulation using the AMST affects the entire aspect of learning or only speed dimension.

The second point to concern relates to the researchers' (Hazan-Liran & Miller, 2017; P. Miller et al., 2018) assumption for the number-copy task. Specifically, with this task it is still uncertain whether the participants remembered all word-number pairs equally or they kept checking specific word-number pairs. For instance, it is possible that the participants did not remember the pair of the word *blue* and its paired number(s) in one or all colour congruency condition(s) whilst they remembered other word-number pairs. If this case happened, it can indicate that although the total number of correct completions was high, the learners might not have remembered all word-number pairs equally. Thus, one might question that their analyses of the "total" number of responses can truly reflect the participants' genuine learning. This limitation seems to hamper from considering that the researchers' (Hazan-Liran & Miller, 2017; P. Miller et al., 2018) findings occurred due to the genuine learning processes.

The final issue is regarding from which process the facilitation effects came from during the task. In other words, there is a possibility that the researchers' (Hazan-Liran & Miller, 2017) observed facilitation effects occurred during searching processes rather than during inputting word-number information. In their study the facilitation was defined by the higher number of completions in the congruent condition compared to the neutral condition. However, this could be due to the direct search strategy to identify the target paired numbers in the example list. For example, in the congruent condition, the participants could directly search the paired number based on its ink colour since only in this condition the ink colours of the numbers (i.e. blue ink printed in the number *7*) matched the meanings of the target colour words (i.e. *blue*). This direct-number search could occur only in the congruent condition, which might have contributed to the highest number of completions in this condition. Since the root of interference was clarified by P. Miller et al.'s (2018) comparison between the incongruent and neutral conditions and their findings (i.e. solving semantic conflict between ink colours and colour words consumed the limited cognitive resources), this can ascertain that the interference effect occurs due to the activation of contradicting information between ink colours and colour words during inputting information rather than during search processes. Therefore, Hazan-Liran and Miller's (2017) combination of the AMST with the number-copy task raises a question in terms of confirming that the facilitation occurs due to learning processes, not due to the visual search strategy.

These limitations would be overcome by removing the pressure on learning fast, imposing pressure on learning all materials equally and removing the processes of searching materials during learning. Thus, it would be important to provide the equal study duration, to set the circumstance that the participants are not informed which materials will be tested and to remove visual search processes. These would be achieved by measuring the task-irrelevant ink colour effects during paired-associate learning with traditional memory tests. There are three reasons for why memory performance is expected to reflect performance of paired-associate learning. First, the definition of learning involves changes in memories (Schnotz & Kürschner, 2007). Second, the reports of the literature (e.g. Hales & Brewer, 2011; Nakazawa et al., 2002; Rajji et al., 2006; Shimamura et al., 1995) indicated that paired-associate learning involves creation of associative memories. Finally, the traditional paired-associate learning research used memory tests to measure its performance (e.g. Paivio, 1965; Paivio & Yarmey, 1966). For these reasons, the present thesis assumes that if task-irrelevant colour-word associations produced in the AMST truly affects learning efficacy, such effects should be also observed on memory tests. Therefore, the current research measures the credibility of the AMST from a perspective of memory performance.

Overview of the Present Research

To recap, the present thesis is based on the following points; (a) the research of paired-associate learning seems to attempt to find a way to optimise the quality of this learning, (b) with the traditional Stroop literature, how simply seeing the task-irrelevant ink colours affects other cognitive performance (i.e. other than colour naming) was unclear, (c) cognitive load theory proposes that dealing with task-irrelevant information modifies performance of learning, (d) Hazan-Liran and colleagues (Hazan-Liran & Miller, 2017; P. Miller et al., 2018) reported that seeing the task-irrelevant ink colours affects the quality of paired-associate learning; however, (e) the limitations of their methodology appears to make their findings questionable whilst these might be overcome by examining the credibility of their paradigm with memory tests.

When designing the memory research, selection of the types of memory tests would be controversial. Since cued-recall tests and associative recognition tests seem to be the common techniques to examine performance of paired-associate learning (Arndt, 2012), the present thesis combines the AMST with these memory tests. Generally, the participants in the present research were asked to remember the pairs of colour concepts and numbers in the form of the AMST. Later, they were asked either to recall the paired numbers cued by the colour concepts or to judge whether the presented pairs were the correct pairings or not (see Figure 1.3).

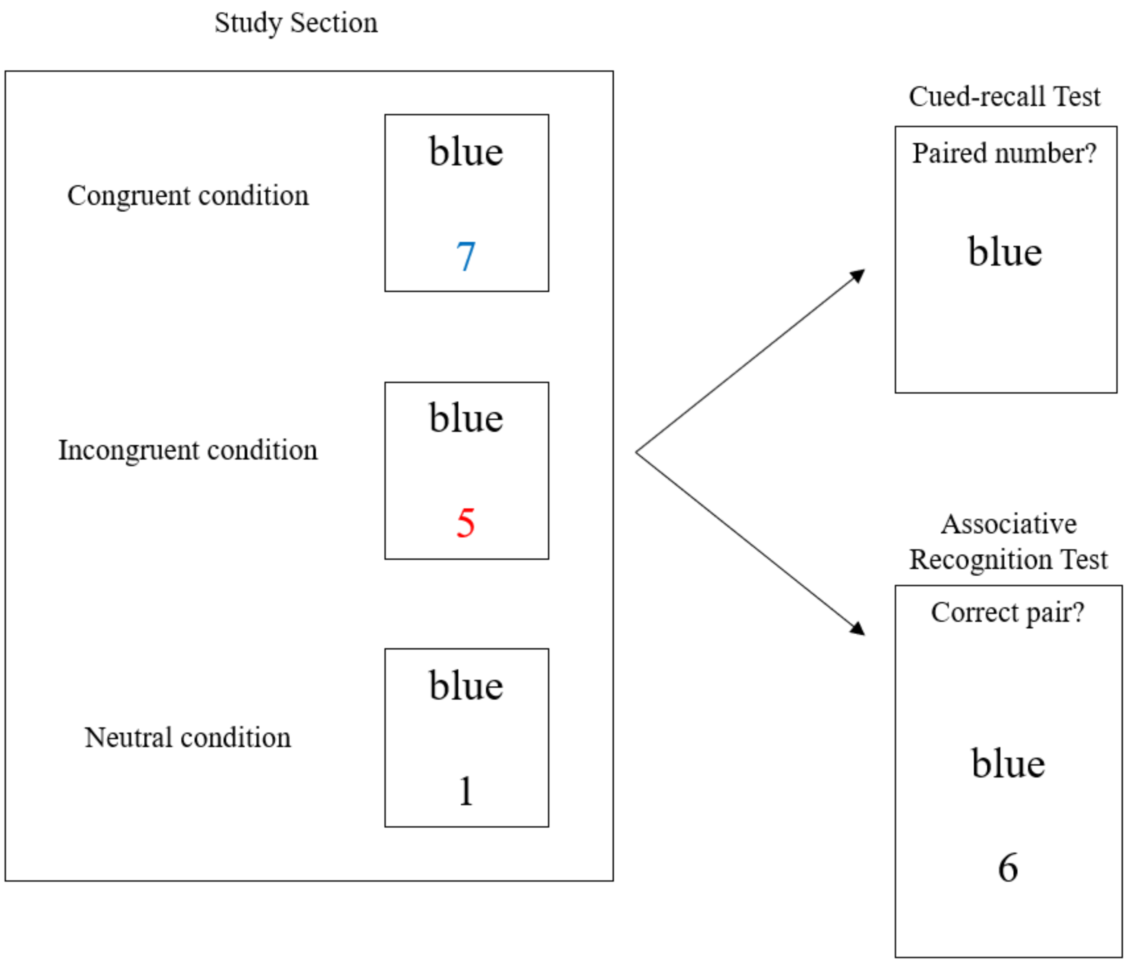


Figure 1.3. The example of the present research procedure. The general procedure of the present thesis was following; (1) the eight pairs of colour concepts and numbers were presented in one of ink colour types whilst each pair was presented for 3 seconds, (2) after studying all eight pairs, they underwent the cued-recall test or the associative recognition test, (3) upon the completion of the memory test section, they started a new ink colour condition that required them to remember a new set of pairs.

The present thesis attempts to answer the research question; whether the AMST can examine the effects of task-irrelevant colour-word associations on efficacy of paired-associate learning. The aim is to examine whether the similar facilitation and interference effects as the two original studies (Hazan-Liran & Miller, 2017; P. Miller et al., 2018) reported can be observed when ink colour effects on the AMST are measured by memory accuracy on cued-recall tests and recognition tests. To achieve this aim, memory accuracy on the cued-recall memory tests and associative recognition tests were analysed. The overall hypothesis is that congruency between an ink colour printed in a number and its semantic relationship with a paired colour word on the AMST would facilitate memory accuracy on a cued-recall test and an associative recognition test compared to the neutral ink colour condition (i.e. facilitation) whilst incongruency between such an ink colour and its relationship with a colour word would interfere with cued-recall and recognition memory accuracy in comparison to the neutral ink colour condition (i.e. interference). Additionally, in some experiments the participants undergo repeated study-test sessions. Thus, the analyses of the ink colour effects across practice would provide additional evidence for how having practice modifies the strength of such effects in the AMST. Whilst the Stroop literature has showed the mixing results of the practice effects on the Stroop effect (for an overview see MacLeod, 1991), memory research might signal for a direction of such effects. Specifically, it was found that when the words are paired with several words (e.g. A-A and A-B), this leads to memory confusion of “what was paired with that word?” (for a summary see Wixted, 2005). If colour-word relationships produce effects not only on encoding but also on retrieval, incongruent ink colours printed in digits might induce memory confusion and this confusion may be accumulated across practice. For example, when the participants are asked to recall the paired number cued by the word *blue*, this cue word might trigger a memory of other number that was printed in blue ink. As the memories of words and ink colours of digits are stored more, the magnitude of such memory interference might increase. Therefore, whilst dissociation of ink colour effects between encoding and retrieval would be out of scope, the present research also examines the effects of practice as a secondary question.

Chapter 2 examined the applicability of the AMST to study memory accuracy on a cued-recall test and an associative recognition test with two types of colour words: colour names (e.g. *blue*, *red*) and colour-related words (e.g. *sky*, *strawberry*). Experiment 1 tested cued-recall memory accuracy of colour names and numbers whilst Experiment 2 assessed it with colour-related words and numbers. Experiment 3 examined recognition memory accuracy of colour names and numbers and Experiment 4 tested it with colour-related words and numbers. Chapter 3 explored the applicability of using colour-related words on the AMST with associative recognition tests only. Experiment 5-A aimed to determine which type of emotional words would be more likely to produce effects on recognition memory accuracy and Experiment 5-B examined such memory accuracy with the one that had shown effects in Experiment 5-A. Since Sutton and Altarriba (2016) reported that emotional words can induce specific colour images, the series of Experiment 5 explored the use of this another type of colour-related words with the AMST. Experiment 6 was a 2-day experiment. This adopted memory consolidation processes via sleep to examine the use of colour-related words on the AMST. Experiment 7 tested whether pre-activation of colour-word associations through priming would induce effects from colour-related words and Experiment 8 compared pictorial stimuli and written words to assess whether presenting colour concepts in a different format would induce effects from colour-related concepts on recognition memory accuracy. Chapter 4 examined whether negative emotions activated by incongruent ink colours contribute to interference in the AMST. This chapter employed only cued-recall tests. In Experiment 9 the valence values (i.e. “good” or “bad”) were assigned to the numbers and this experiment tested whether seeing the good/bad numbers during study would modify the interference effects from incongruent ink colours on correctness of responses on the subsequent cued-recall memory tests. Experiment 10 inserted three types of face icons (i.e. happy, sad and neutral) between each word-number pair and examined whether this emotional intervention via face icons would change the interference effect in the incongruent condition. Finally, Experiment 11 replaced the face icons with more influential emotional stimuli (i.e. sound stimuli; positive, negative and neutral) to test whether such robust stimuli could modify the interference effects from incongruent ink colours in the AMST.

Chapter 2: Colour Names and Colour-Related Words on the AMST with Cued-Recall Tests and Recognition Tests

This chapter explores the applicability of the AMST to study learning when ink colour effects on the AMST are examined by two types of memories: cued-recall memory and recognition memory. By comparing performance with two memory tasks, it will confirm that ink colour congruency on the AMST is limited to a specific type of memory or not.

Whilst the original studies (Hazan-Liran & Miller, 2017; P. Miller et al., 2018) demonstrated effects of task-irrelevant ink colours on learning word-number pairs, their conclusion seems to be questionable since their results might have occurred due to visual search strategies, not due to learning (i.e. changes in memories; for a review see Schnotz & Kürschner, 2007). To solve this issue, Chapter 2 combines the AMST with a cued-recall test (Experiment 1 and 2) and an associative recognition task (Experiment 3 and 4), which removed visual search processes during learning. If the present research shows different results from the two studies (Hazan-Liran & Miller, 2017; P. Miller et al., 2018), it would indicate that their findings might have occurred only during visual search processes, not when learners acquire new information. Therefore, Chapter 2 examines the credibility of the AMST as a new tool to measure the effects of task-irrelevant ink colours on one’s learning efficacy by analysing performance on memory tests.

The overall research questions in Chapter 2 were (1) whether it is possible to replicate the findings in the original literature (Hazan-Liran & Miller, 2017; P. Miller et al., 2018) when the AMST is used to study memory accuracy, (2) if so, what type of effects (i.e. facilitation and/or interference) occurs, (3) whether such effects occur with both colour names and colour-related words and (4) whether such effects are affected by the type of memory tests. Particularly, the present chapter aimed to establish the credibility of the AMST as a new paradigm to measure the effects of task-irrelevant ink colours on learning efficacy by extending the findings of the two original studies (Hazan-Liran & Miller, 2017; P. Miller et al., 2018). To achieve this aim, the AMST was combined with cued-recall tests and associative recognition tests. These two memory tests were chosen since Arndt (2012) introduced these two memory tests as typical measurement methods to examine the quality of paired-associate learning. For instance, when the *dog-cat* pair and the *bird-wing* pair are the to-be-remembered pairs, cued-recall tests ask learners to recall the paired item (i.e. *dog*) when one studied item (i.e. *cat*) is presented and associative recognition tests ask them to judge whether the presented item pairs are the correct pairing or the incorrect ones (i.e. *dog-wing*). Note that in associative recognition tests both words are studied previously but pairing might be either correct or incorrect. In the present chapter the main research procedure was that the participants were asked to remember eight word-number pairs for the later cued-recall memory tests (Experiment 1 and 2) or associative recognition memory tests (Experiment 3 and 4) and accuracy of performance on these memory tests was analysed. The overall hypothesis was that the congruent condition would reveal the most superior memory accuracy whilst the incongruent condition would show the poorest memory accuracy amongst three colour congruency conditions (i.e. the congruent, incongruent and neutral conditions) on both memory tests.

Experiment 1

The first experiment examined whether cued-recall memory tests can reflect the effects of task-irrelevant ink colours on the AMST when the words are colour names (e.g. *red*, *blue*). As Arndt (2012) noted, the quality of paired-associate learning can be measured on a cued-recall test or an associative recognition test. Since the number-copy task in the two original studies (Hazan-Liran & Miller, 2017; P. Miller et al., 2018) appears to involve how fast and accurately the participants can recall appropriate information, a recall test would be appropriate as a starting point of the present research.

There are various ways to measure recall memory performance. Generally, a recall task can be decomposed into three types: a free recall task, serial recall task and cued-recall task (Nieuwenhuis-Mark, 2012). A free recall task requires participants to recall studied items regardless of its order whilst a serial recall test asks them to recall such information in the same order that they remembered (Golomb, Peelle, Addis, Kahana, & Wingfield, 2008). According to Arndt's (2012) explanation, on a cued-recall test, participants first see pairs of items (e.g. *stove*-*letter*). After that, one member of pairs (e.g. *stove*) is presented and participants are asked to recall its paired item (e.g. *letter*). Therefore, whilst there are three common techniques (i.e. a free recall task, serial recall task and cued-recall task) to measure recall memory, the present research focuses on a cued-recall task since this paradigm seems to be the closest to the original methodology (i.e. recalling the paired numbers that were cued by colour words) in the two studies (Hazan-Liran & Miller, 2017; P. Miller et al., 2018).

The task-irrelevant ink colour congruency on the AMST was expected to affect cued-recall memory performance due to enhancement of encoding and of recovery by additional information. The former reason was suggested based on the cued-recall memory enhancement by synesthetic colours (i.e. the automatically activated colours by people who involuntarily perceive additional sensation such as seeing colours in numbers) in written words (Mills, Innis, Westendorf, Owsianiecki, & McDonald, 2006); seeing synesthetic colours (i.e. additional information) in words helped synesthetes to remember the written words accurately. The latter was argued since semantically-related contextual cues during cued-recall were reported to have facilitative effects (Roome, Towse, & Crespo-Llado, 2019). Roome et al. described such effects in the framework of the reconstruction hypothesis, which assumes that additional information (e.g. contextual cues) recover or enhance contents in working memory (Cowan et al., 2003; Towse, Cowan, Hitch, & Horton, 2008; Towse, Hitch, Horton, & Harvey, 2010). When applying these findings to the AMST and memory, it would lead to an assumption that ink colours printed in numbers (i.e. contextual information) might affect cued-recall memory performance. On the AMST, colour words (e.g. *blue*) are written in black ink and the paired numbers are printed in either congruent (e.g. blue ink), incongruent (e.g. brown ink) or neutral ink colours (e.g. black ink). When the numbers are coloured, such additional contextual information may help learners' encoding of the numbers; thus, memory performance between the (in)congruent and neutral conditions was expected to be different. Specifically, the semantically-related contextual cues would occur only in the congruent condition since incongruent and neutral ink colours are semantically unrelated to colour names. Rather, as Hazan-Liran and colleagues (Hazan-Liran & Miller, 2017; P. Miller et al., 2018) demonstrated, processing incongruent ink colours might produce harmful effects on memory performance due to the less amount of available cognitive resources to remember word-number pairs.

However, it is important to consider the risk of memory impairment due to natural difficulty of recall memory tests. When comparing memory tests, a recall task tends to be more demanding than recognition tests since learners have to recall the target information precisely (Haist, Shimamura, & Squire, 1992). In the framework of the cognitive load theory, Sweller (2010) explained that when more working memory resources are used to deal with a difficult task (i.e. high intrinsic cognitive load), less resources are left to process or manage task-irrelevant information (i.e. extraneous cognitive load). Therefore, if an assigned task is too demanding for participants, their cognitive resources would be absorbed by task difficulty only. That is, in the present research, there would be a risk that participants' task performance with three types of colour congruency (i.e. congruent, incongruent and neutral) were equally impaired due to difficulty of the recall tests; consequently, there might be no memory performance difference between these three types of ink colours. In order to study the effects of task-irrelevant ink colours on memory, this situation needs to be avoided. To prevent it, increasing learners' familiarity with the recall tests would be necessary. For this reason, the present research increases opportunities of studying and testing word-number pairs. The review article of the Stroop effect (MacLeod, 1991) illustrated the mixing effects of practice on the magnitude of this effect. Therefore, the present study will also demonstrate additional evidence for whether increasing learners' task familiarity modifies the degree of colour congruency effects on the AMST as the secondary question.

The research question of Experiment 1 was whether it is possible to replicate the findings of the two original studies (Hazan-Liran & Miller, 2017; P. Miller et al., 2018) when the effects of task-irrelevant ink colours on the AMST is measured by cued-recall tests and the words on the AMST are colour names. The aim of Experiment 1 was to examine whether cued-recall memory accuracy for a word-number pair is influenced by an irrelevant ink colour printed with the number and its relationship to the word. To achieve this aim, Experiment 1 asked the participants to remember eight pairs of colour names and numbers with the colour names were written in black ink and the numbers were presented in other ink colours (e.g. red ink). Later, the participants were asked to type the correct paired numbers when the colour names appeared on the screen one by one. The numbers of correct and incorrect responses were recorded and their reaction times in each trial were also recorded to examine the effects of speed-accuracy tradeoff (i.e. whether participants give priority to responding fast at the cost of responding accurately or vice versa; Zimmerman, 2011). The hypothesis in Experiment 1 was that the congruent condition would show the highest cued-recall test scores whilst the incongruent condition would reveal the lowest cued-recall test scores in all study-test sessions. For practice effects, the memory literature (e.g. Wixted, 2005) has reported that associating the one stimulus with multiple concepts may lead to memory confusion. Thus, it suggested the increment of the ink colour effects if ink colours produce effects not only on encoding but also on retrieval. However, it is uncertain that such a practice effect is applicable to the memory version of the Stroop task. Therefore, the secondary hypothesis in the present experiment was that the degree of ink colours would be increased as the number of study-test sessions increase.

Method

Participants

Thirty students of University of Kent at Canterbury were recruited as the participants. In the present thesis all participants participated in the experiments for course credits with the prerequisite of having normal or corrected eye vision. The School of Psychology Ethics committee at the University of Kent provided ethical approval for all experiments that conformed to the British Psychological Society code of ethics. The sample size was determined based on the number of participants used by Hazan-Liran and Miller (2017). Additionally, Gpower was used to calculate the sample size based on the performance scores reported by Hazan-Liran and Miller (2017) Table 1. According to this analysis, a sample of 10 was required to achieve a power of .80 for both interference and facilitation; therefore, 30 participants were recruited. Five male participants (M = 18.60 years old; SD = 0.89) and 25 female participants (M = 19.28 years old; SD = 1.60) were recruited and these 30 students consisted of 26 native English speakers, three non-native English speakers and one "English plus other language(s)" speaker.

Design

Experiment 1 mixed the AMST from the two original studies (Hazan-Liran & Miller, 2017; P. Miller et al., 2018) with a cued-recall test, which asked the participants to recall and type the correct paired numbers when the colour names appeared on the screen. There were three colour congruency conditions: the congruent (e.g. *blue*-*7* with the word *blue* in black ink and the number *7* in blue ink), incongruent (e.g. *blue*-*5* with the word *blue* in black ink and the number *5* in brown ink) and neutral (e.g. *blue*-*1* pair with both word and number in black ink) conditions. To ascertain that the participants became familiar with the task, the participants repeated study-test sessions three times in each colour congruency condition. This number of study-test session repetition was determined arbitrarily. Thus, a two factorial design was used with Colour congruency (Congruent, Incongruent and Neutral) and Study-test session (1st, 2nd and 3rd) as within-participants factors. The independent variable was colour congruency and the dependent variable was the proportion of correct responses. The reaction time was also recorded. The order of colour congruency conditions was counterbalanced based on Williams Latin Square Design (Williams, 1949).

Materials

All experiments in the present thesis were conducted with the software Psychopy (Peirce et al., 2019). The settings of words and numbers were identical to the two original studies (Hazan-Liran & Miller, 2017; P. Miller et al., 2018) except these stimuli were written in English spellings and presented in Arial font and in the size of 0.2 in Psychopy with a grey background.

The word frequency per million word tokens in English was obtained from *Word Frequencies in Written and Spoken English: Based on the British National Corpus* (Leech, Rayson, & Wilson, 2001). The eight colour names (Word frequency; [RGB codes]) are listed here in the alphabetical order: *blue* (92; [0, 0, 255]), *brown* (45; [165, 42, 42]), *green* (101; [0, 128, 0]), *orange* (14; [255, 165, 0]), *purple* (11; [128, 0, 128]), *red* (126; [255, 0, 0]), *white* (207; [255, 255, 255]) and *yellow* (41; [255, 255, 0]). Their average frequency was approximately 79.63 per million word tokens and the raw average letter length was 5.00 letters. The seven colour names (*blue*, *brown*, *green*, *orange*, *red*, *white* and *yellow*) were obtained from the two original studies (Hazan-Liran & Miller, 2017; P. Miller et al., 2018) but the word *black* was replaced with the word *purple* in the present thesis. This replacement was performed since black ink was used in the neutral condition. For example, the incongruent condition needed to present colour names and numbers with incongruent ink colours (e.g. *blue*-*4* with the word *blue* in black ink but the number *4* in red ink). However, if both words and numbers were written in black ink (*blue*-*4* pair with both word and number in black ink), it would be difficult to see black ink as incongruent ink colours; rather, the participants would be more likely to see black ink as a neutral ink colour. In this case, such pairs might fail to produce semantic conflict and to examine ink colour congruency effects. To avoid this risk, the present thesis avoided to use the word *black* in all colour congruency conditions and black ink in the two colour congruency conditions. These stimuli were replaced with the word *purple* and purple ink. Therefore, all eight words were written in black ink but the numbers were presented in other ink colours (e.g. red ink) in the two colour congruency conditions whilst both words and numbers were written in black ink in the neutral condition. The eight numbers were digits from *1* to *8* in Arabic format. These words and numbers were presented at the centre of the screen; all words were presented in the horizontal distance value (i.e. the horizontal distance from the centre of the screen) of 0 and the vertical distance value (i.e. the vertical distance from the centre of the screen) of 0.2 position. The numbers were set in the horizontal distance value of 0 and the vertical distance value of -0.2 position in Psychopy. Simply, the words and numbers were presented at the centre of the screen whilst the words were presented above the numbers.

Experiment 1 presented all combinations of colour names and incongruent ink colours rather than selecting incongruent ink colours randomly. In the incongruent condition the meanings of the colour names and the ink colours of the numbers needed to be incongruent (e.g. presenting the *blue*-*6* pair with red ink). When presenting the word-number pairs in the incongruent condition, there were seven possible combinations of colour names and incongruent ink colours (see Appendix A). For example, when the participants saw the colour name *blue*, the incongruent ink colour could be one of other seven ink colours: brown, green, orange, purple, red, white or yellow ink. This experiment included all of seven patterns and selected one of the seven patterns to each participant. The order of these seven incongruent colour patterns was rotated. That is, Participant 1 experienced Pattern 1 (e.g. seeing the *blue*-*6* pair with the number *6* in brown ink), Participant 2 saw Pattern 2 (e.g. seeing the *blue*-*6* pair with the number *6* in yellow ink) and so forth. Thus, the combinations of colour names and incongruent ink colours were different depending on Participant numbers. It was possible for the software to select incongruent ink colours randomly. However, Experiment 1 did not employ this random selection in order to eliminate the potential risk of unequal frequency of ink colours. If the software randomly chose incongruent ink colours, there would be a possibility that some ink colours were not selected as incongruent ink colours. For example, the participants might see the *blue*-*6* pair with the number *6* in red ink whilst they might see all other seven word-number pairs (i.e. *brown*-*5*, *green*-*3*, *orange*-*7*, *purple*-*2*, *red*-*8*, *white*-*4* and *yellow*-*1*) with the numbers in blue ink. In this case, it would be difficult to examine how other six ink colours (i.e. brown ink, green ink, orange ink, purple ink, white ink and yellow ink) influence on memory performance as incongruent ink colours. That is, since only red ink and blue ink were used as incongruent ink colours, there might be a possibility of presenting the *blue*-*6* pair with red ink differs from presenting it with brown ink. Thus, presenting all combination patterns of colour names and incongruent ink colours (e.g. presenting the *blue*-*6* pair with brown, green, orange, purple, red, white and yellow inks) can reduce this risk and the results can include the effects of all ink colours as incongruent ink colours. Therefore, Experiment 1 showed all seven combinations of colour names and incongruent ink colours rather than using the random selection of the software.

The participants studied the same word-number pairings during three study-test sessions in each colour congruency condition. However, each colour congruency condition presented different lists of the eight word-number pairs (e.g. the participants saw the *blue*-*4* pair in the congruent condition but the *blue*-*6* pair in the incongruent condition; see Appendix A). The words and numbers were randomly paired. In the present thesis all randomisation processes were performed via the website (Randomness and Integrity Services Ltd, n.d.). The same word-number pairings were used across all participants; that is, all participants saw the *blue*-*4* pair in the congruent condition.

To reduce the carry-over effects from the previous condition, short tone tasks were inserted between each colour congruency condition. The tones were downloaded from the two websites. The 100 Hz, 250 Hz, 440 Hz and 1000 Hz frequency tones (under the category of "Duration 0:05") were downloaded from the website ("MediaCollege.com - Download Audio Tone Files," n.d.). These four tones were chosen because these are the first four tones in the list of the website. The 340 Hz, 660 Hz, 800 Hz and 2400 Hz frequency tones were downloaded from another website ("BEKA - BR385 tone samples," 2017). These four tones were selected since these are also the first four continuous tones in the list.

Procedure

In all experiments of the present thesis all participants were tested individually in a quiet lab with a computer and all experiments were advertised as the study of learning. The experimenter did not particularly mention regarding ink colours as the two original studies (Hazan-Liran & Miller, 2017; P. Miller et al., 2018) did not provide such information to the participants. All experiments in the present thesis were advertised as a study of memory that asked the participants to learn pairs of items. The written informed consent explained that they could withdraw at any time for any reasons. Only after the participants signed the informed consent in front of the computer, the experiment started and the experimenter asked their demographic information (i.e. age, gender and native language). After the experimenter left the room, the participants saw the instructions about the experiment in black ink with a grey background. The screen informed that the task was to remember the presented word-number (digits from *1* to *8*) pairs and later they would be asked to type the paired numbers as quickly and accurately as possible when the words appeared on the screen.

The present experiment used the top row of a QWERTY keyboard. In the beginning of Experiment 1 the screen instructed the participants to set their left little finger on the “1” key, left ring finger on “2,” left middle finger on “3,” left index finger on “4,” right index finger on “5,” right middle finger on “6,” right ring finger on “7” and right little finger on “8.” On each trial during the practice and test sections, the picture of the hands (“typingMe.com,” n.d.) that presented the assigned finger placement was shown permanently on the screen as a reminder. Note that since the original picture was the simple picture of hands, the numbers printed in black were added to the original picture for the present research. After the instruction, the participants experienced the 48-number-typing practice trials with feedbacks. During these practice trials, the participants saw the numbers (digits from *1* to *8*) on the screen and were asked to type the presented numbers with the assigned finger placement. These numbers were kept visible on screen until responses were made. When the participants pressed the correct number keys, the screen presented the feedback message “Correct!” whilst the participants saw the feedback message “Oops! That was wrong” when they typed the incorrect number keys.

These feedback messages were presented for 500 milliseconds (ms). Upon the completion of the 48-number-typing practice trials, the screen explained that in the study section the participants would be asked to remember eight word-number pairs and in the test section they would need to type the paired numbers as quickly and accurately as possible when the words appeared on the screen. The screen also informed that there would be three study-test sessions and these three study-test sessions would repeat three times (i.e. 3 study-test sessions x 3 colour congruency conditions). Note that the participants were informed that they would study the same word-number pairs across three study-test sessions. The instruction explained that the participants would have 2 minutes to complete eight test trials. This time limit was set to ascertain participants' appropriate focus on the task.

In the study section the participants saw eight word-number pairs one at a time, which each pair was presented for 3,000 ms. The order of the eight word-number pairs were randomised for each participant by the software. The words were presented in black ink whilst the numbers were written in either the congruent ink colours, the incongruent ink colours or black ink. When the participants finished seeing all eight word-number pairs, a blank screen appeared for 500 ms between the study section and the test section in order to prevent the same word from appearing continuously. When this blank screen disappeared, the test section started and the participants were required to type the correct paired numbers when the colour names appeared on the screen. In the test section the colour names were written in black ink and the participants' responses were invisible on the screen. The colour names appeared on the screen one at a time and these words were kept visible until the participants made responses. The order of these words was also randomised for each participant by the software. The test section automatically ended when the participants completed all eight test trials or 2 minutes passed. This study-test session repeated three times with the same word-number pairs. When the participants finished the third study-test session, the 16-trial tone comparison task was inserted and after this, a new ink colour condition began; the screen informed that the participants would need to study a new set of eight word-number pairs.

Between each colour congruency condition, the 16-trial tone comparison task was inserted to reduce the carry-over effects of the previous condition. Two pure tones were presented sequentially with each exposure duration of 500 ms and a blank interval (for 500 ms) between the two tones. Additionally, 1,500-ms inter-trial interval was inserted. The two tones repeated until the participants judged whether two sequentially presented tones were the same or different with using the “C” key (with the left index finger) for the same tones and the “I” key (with the right index finger) for the different tones. There were eight “same” tone pairs (e.g. 100 Hz-100 Hz) and eight “different” tone pairs (e.g. 250 Hz-340 Hz). Since the participants studied eight word-number pairs, this tone comparison task consisted of the eight “same” tone pairs and the eight “different” tone pairs. This 16-trial tone comparison task also automatically ended for completion of all 16 trials or passing 2 minutes. The orders of all tone pairs were also randomised for each participant. Therefore, there were the total of 72 test trials (8 word-number pairs x 3 study-test sessions x 3 colour congruency conditions) and the two 16-trial tone comparison tasks. This experiment took approximately 15 minutes to complete.

Results

All analyses in the present thesis were performed in IBM SPSS Statistics Version 25 and was based on the alpha level of .05. Additionally, when Mauchly’s Test of Sphericity was violated in the current thesis, the Greenhouse-Geisser corrected values were reported. When an analysis of variance (ANOVA) was used in the present research, Bonferroni correction was used. For the 2-minute time limit, none of the participants was timed out. Three participants made responses in less than 200 ms. However, each of them made such responses only in one trial, two trials and three trials out of 72 test trials, respectively, whilst all participants' mean reaction times in each study-test session of all colour congruency conditions were longer than 200 ms. Since these could indicate that all of them paid appropriate attention to the task, all responses were included in the following analyses.

Analysis of Accuracy

Table 2.1 and Figure 2.1 showed a summary of descriptive statistics. First, to examine whether colour-word congruency produced any different memory performance, only the congruent and incongruent conditions were compared. The 2 (Colour congruency) x 3 (Study-Test Session) within-participants factorial ANOVA was conducted and revealed the significant main effect of Colour congruency [$F(1, 29) = 17.90$, $p < .001$, partial-$\eta^2 = .38$] and of Study-test session [$F(1.604, 46.519) = 46.05$, $p < .001$, partial-$\eta^2 = .61$]. The analysis revealed no interaction effect of Colour congruency x Study-Test session [$F(2, 58) = 2.61$, $p = .082$, partial-$\eta^2 = .08$]. This indicated that memory accuracy between the congruent and incongruent conditions were significantly different but the degree of this difference was identical across the three sessions.

To know the direction of ink colour effects (i.e. facilitation and interference) and examine if the magnitude of these effects were affected by the number of sessions, the neutral condition was included. The proportion correct scores were analysed in a 3 (Colour Congruency) x 3 (Study-Test Session) within-participants factorial ANOVA.

Table 2.1

Mean Score Proportions, Reaction Times (in ms) and Mean Incorrect Response Frequency Proportions with Three Study-Test Sessions in Each Colour Congruency Condition (Experiment 1)

Condition	Study-Test Session	Accuracy	Reaction Time	"Same" Incorrect Response Frequency Proportion	"Not_same" Incorrect Response Frequency Proportion
		M (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)
Congruent	1st	.37 (.27)	2,282 (694)	—	—
	2nd	.65 (.23)	2,454	—	—

			(1,132)		
	3rd	.75 (.30)	2,153 (814)	—	—
Incongruent	1st	.24 (.14)	2,365 (677)	.15 (.15)	.85 (.15)
	2nd	.39 (.27)	2,449 (959)	.14 (.15)	.79 (.26)
	3rd	.54 (.30)	2,357 (839)	.24 (.25)	.66 (.33)
Neutral	1st	.34 (.22)	2,374 (834)	—	—
	2nd	.53 (.25)	2,460 (985)	—	—
	3rd	.66 (.26)	2,358 (919)	—	—

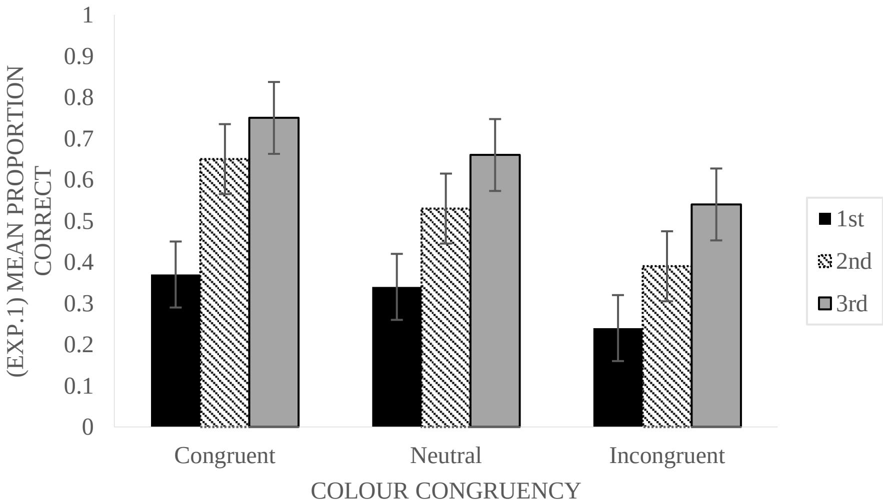


Figure 2.1. The graph shows the mean proportions of correct scores in each colour congruency condition (Experiment 1). Error bars represent the 95% confidence intervals (Masson & Loftus, 2003).

In Experiment 1 there was a significant main effect of Colour Congruency, $F(2, 58) = 9.33$, $p < .001$, partial-$\eta^2 = .24$. Paired sample *t*-tests (two-tailed) showed that the incongruent condition ($M = .39$, $SD = .20$) was less accurate than the neutral condition ($M = .51$, $SD = .19$), $t(29) = 2.69$, $p =.012$, $d = 0.49$, and the congruent condition ($M = .59$, $SD = .22$), $t(29) = 4.23$, $p < .001$, $d = 0.77$. There was no significant difference between the congruent condition and the neutral condition, $t(29) = 1.66$, $p = .107$, $d = 0.31$. There was a main effect of Study-Test Session, $F(2, 58) = 61.38$, $p < .001$, partial-$\eta^2 = .68$. This indicated that the participants' scores significantly increased with repetition; from the first session to the second session [$t(29) = 7.99$, $p < .001$, $d = 1.46$] and from the second to the third study test session [$t(29) = 4.29$, $p < .001$, $d = 0.78$]. The interaction effect of Colour Congruency x Study-Test Session was not significant, $F(4, 116) = 1.31$, $p = .270$, partial-$\eta^2 = .04$. The analyses of reaction times confirmed no occurrence of speed-accuracy tradeoff (all $p > .05$). Speed-accuracy tradeoff was also examined in other experiments; however, all analyses indicated no occurrence of speed-accuracy tradeoff in the present research. Therefore, the analyses of reaction times will not be reported in the subsequent experiments.

Whilst the main effect of Colour congruency indicated no facilitation, Figure 2.1 seems to illustrate a possibility of the occurrence of facilitative effects with sessions 2 and 3. Therefore, the same 3 x 3 ANOVA was performed without session 1, which resulted in a significant main effect of Colour congruency and an insignificant interaction effect of Colour congruency and Study-Test Session. By paired sample *t*-test (one-tailed), the facilitation effect was confirmed, $t(29) = 1.92$, $p = .032$, $d = 0.35$.

Since there were six patterns of condition order (e.g. congruent-incongruent-neutral, incongruent-neutral-congruent), the 3 (Colour) x 3 (Session) x 6 (Condition Order) factorial ANOVA was conducted to examine whether the condition order had any effects. This analysis showed the interaction effect of Colour x Repetition x Condition Order was not significant [$F(20, 96) = 0.93$, $p = .548$, partial-$\eta^2 = .16$]. Consequently, this analysis seemed to suggest that the condition order had little impact in the present experiment. This effect of the condition order was also examined in other experiments; however, no interactions were found with the condition order on memory accuracy. Therefore, the effect of this factor will not be reported in the subsequent experiments.

Since there were seven patterns of incongruent colour combinations as a between-participant factor, this was entered in the 3 (Colour Congruency) x 3 (Study-Test Session) factorial ANOVA to confirm that seeing the *blue*-*6* pair with brown ink and yellow ink produced the same results. The analyses confirmed this; the interaction effects with incongruent colour combination patterns were all insignificant (all $p > .05$). Additionally, there was no memory accuracy difference between these seven groups, $F(6, 23) = 0.39$, $p = .877$, partial-$\eta^2 = .09$.

For the poorest accuracy in the incongruent condition, one might question whether the intrinsic property from the previous trials might have increased the confusion and led to the poorest learning. In other words, interference might have occurred because the ink colours or colour names in the previous trial affected accuracy in the following trials. This memory interaction between newly input information and past memories has been reported by the literature of interference theory (e.g. Anderson, 2003; Ankala, 2011; Ceraso, 1967). In the present experiment some participants saw the *yellow-1* pair with the number *1* was printed in blue ink. Since they saw the word *yellow*, the colour image of “yellow” would be activated and held in their working memory. However, immediately after this pair, the *white-4* with the number *4* was written in yellow ink could be presented due to the randomisation of the word-number pair order. In this case, whilst they were holding the colour image of “yellow,” they saw the actual yellow ink although the source of these two types of information was different word-number pairs (i.e. *yellow-1* and *white-4*). The similar situation could occur with ink colours to words. For example, the participants saw the same *yellow-1* pair, the colour image of “blue” also remained active since the ink colour of the number *1* was blue. When they saw the *blue-6* (the number *6* was printed in brown) pair immediately after the *yellow-1* pair, the same colour image (i.e. the colour image of “blue”) could be also activated due to seeing the word *blue*. Thus, the same working memory content was triggered twice from the different word-number pairs, which might have accidentally provided a link between the two different word-number pairs. Therefore, it could be possible to assume that this accidental matching of activated colour images and actual visual stimuli might have increased the difficulty of learning association between word-number pairs and resulted in poor memory performance. To calculate whether the frequency of this previous-trial overlap is related to memory accuracy, the analysis involved two factors: Word shift and Colour shift. Word shift represents a case when the participants saw the ink colours that the colour names in the previous trials refer to (e.g. seeing the *yellow-1* pair and then seeing the *white-4* with yellow ink). Colour shift is a situation where they saw the colour names that was activated by the ink colours in the previous trials (e.g. seeing the *yellow-1* pair with blue ink and then seeing the *blue-6* pair). The number of these cases happened was calculated in each session and it was divided by 7 (since there were eight word-number pairs in each session and the first word-number pair was not affected by these two factors) in each session. Therefore, the proportion of the frequency of Word shift and of Colour shift were entered simultaneously as predictors of memory accuracy in the first, second and third study-test sessions in multiple regression. The analysis revealed that there was no significant regression equation between these two predictors (Word shift proportion in the first session: $\beta = -.22$, $p = .256$; Colour shift proportion in the first session: $\beta = .02$, $p = .900$) and memory performance in the first session [$F(2, 27) = 0.74$, $p = .489$], which two variables together explained only 5.2% of the variance. It also showed the identical insignificant regression in the second session (Word shift proportion in the second session: $\beta = .002$, $p = .993$; Colour shift proportion in the second session: $\beta = -.31$, $p = .103$), $F(2, 27) = 1.43$, $p = .256$, which two variables together explained only 9.6% of the variance. Finally, again these two variables (Word shift proportion in the third session: $\beta = .11$, $p = .577$; Colour shift proportion in the third session: $\beta = -.03$, $p = .876$) did not significantly predict memory performance in the third session [$F(2, 27) = 0.19$, $p = .828$], which two variables together explained only 1.4% of the variance. Therefore, these results of the present experiment indicated that the accidental matching between the activated working memory contents from the previous trials and from the current to-be-processed stimuli have little impact on memory accuracy with the AMST.

Regarding the memory interaction between the past and new memories, there was another point to analyse: the pattern of incorrect responses in the incongruent condition. Particularly, when the participants typed the incorrect numbers as a response to the presented colour names, it would be possible that they typed the incorrect numbers that had been printed in the ink colours, which the presented colour names refer to. For example, some participants saw the *yellow-1* pair with the number *1* was printed in blue ink and the *white-4* with the number *4* was written in yellow ink during study. In the test section when the word *yellow* appeared, the correct response was to press the “1” key. However, since they saw the word *yellow* and the number *4* had been written in yellow ink, it could be possible that the activation of colour image “yellow” from seeing the word *yellow* triggered the number memory of *4*; consequently, the memory performance was the poorest in the incongruent condition. If the number of this type of incorrect responses occurred frequently, this could indicate that the participants made mistakes not randomly. This scenario was examined by focusing on the participants’ incorrect responses to clarify which ink colours had been used to present these incorrectly-typed numbers during study. When the test colour names and the previous ink colours of their incorrect responses referred to the same meanings (e.g. seeing the word *yellow* and typing the incorrect number *4*, which had been presented in yellow ink), this situation was named “Same” whilst the disagreement between these two factors (e.g. seeing the word *yellow* and typing the incorrect number *8*, which had been presented in white ink) was entered as “Not_same.” The frequency of these two factors in each session was calculated and divided by the total number of incorrect responses in each session (see Table 2.1).

These response proportions of Same and Not_same were compared via the 3 (Study-Test Session) x 2 (Response) factorial ANOVA with Study-Test Session (i.e. first, second, third) and Response (i.e. Same and Not_same) as within-participant factors. Whilst the main effect of Study-Test Session was not significant [$F(2, 58) = 1.80$, $p = .175$, partial-$\eta^2 = .06$], there was the significant main effect of Response [$F(1, 29) = 188.52$, $p < .001$, partial-$\eta^2 = .87$] as well as the significant interaction effect of Study-Test Session x Response [$F(2, 58)$

= 5.12, p = .009, partial-η^2 = .15]. This indicated that the proportion of Not_same frequency (M = .82, SD = .11) was significantly higher than of Same (M = .18, SD = .11) throughout the experiment whilst the robustness of this difference was changed depending on the number of study-test sessions. For the interaction effect, paired sample t-tests (two-tailed) revealed that, from the first session to the second session, the proportions of Same frequency did not significantly change [Same (1st): M = .15, SD = .15; Same (2nd): M = .14, SD = .15], t(29) = 0.14, p = .894, d = 0.02. This pattern also occurred with the proportions of Not_same frequency [Not_same (1st): M = .86, SD = .15; Not_same (2nd): M = .79, SD = .26], t(29) = 1.19, p = .243, d = 0.22. However, between the second to the third session, the proportion of Same frequency significantly increased [Same (3rd): M = .24, SD = .25], t(29) = -2.17, p = .038, d = 0.40, whilst the proportion of Not_same frequency did not [Not_same (3rd): M = .66, SD = .33], t(29) = 1.82, p = .079, d = 0.33. In other words, when the participants saw the word *yellow*, the chance of typing the incorrect number *4* (which had been presented in yellow ink) significantly increased in the third session compared to the first and second sessions. This could indicate that in the first and second sessions, when the participants were trying to recall what the paired number was, the recalled information was unlikely to include ink colour information. However, with the experience of more frequent study opportunities, the recalled memories might have started including ink colour information, which might have led the participants to make the "Same" type of incorrect responses more often than the first two sessions. Whilst this type of incorrect responses showed a relationship with the number of study-test sessions, the frequency of these was significantly lower than of the Not_same responses. Thus, this would indicate little impact of the Same type responses on memory performance in the present experiment.

Discussion

Overall, Experiment 1 found that the participants remembered the pairs of colour names and numbers less accurately when the numbers were written in the incongruent ink colours (compared to in the congruent ink colours and in black ink), in the second and third study-test sessions the congruent ink colours produced memory enhancement and the magnitude of these effects was not interacting with the number of study-test sessions. To extend the applicability of the findings of Experiment 1, it is necessary to examine whether the identical ink colour effect can occur with colour-related words (e.g. *sky*, *lemon*) as the original studies (Hazan-Liran & Miller, 2017; P. Miller et al., 2018) examined. Therefore, Experiment 2 tested whether the cued-recall memory tests can reflect the effects of task-irrelevant ink colours on the AMST when the words on the AMST are colour-related words.

Experiment 2

Experiment 2 examined whether the same pattern of the results as Experiment 1 would be observed when the words on the AMST are colour-related words. In the present thesis colour-related words refer to words that induce specific colour images such as the word *sky*, *lemon* or *strawberry*. The two studies conducted by Hazan-Liran and Miller (2017) and P. Miller, Hazan-Liran and Cohen (2018) reported effects with colour-related words whilst the magnitude of these effects were weaker than when testing with colour names.

Such waning of the degree of the effects were also expected in the present experiment by considering the directness of semantic links. The literature of the Stroop effect (Levin & Tzelgov, 2016) described the underlying model for why the relatedness of semantic associations between colours and colour words affect the magnitude of the Stroop-related effects. It has been reported that reading colour names (e.g. *red*) “directly” activate its associated colour concepts (e.g. the colour concept of “red”) whereas colour-related words (e.g. *tomato*) “indirectly” trigger those concepts and when concepts are indirectly linked, the activation of subsequent concepts becomes weaker compared to when concepts have a direct semantic relationship based on the model of the semantic network model (Collins & Loftus, 1975). Combining these, the literature (Levin & Tzelgov, 2016) explained that colour-related words tend to produce the weaker ink colour congruency effects compared to colour names.

Therefore, the research question in Experiment 2 was whether the cued-recall memory tests can reflect the effects of weak ink colour congruency on the AMST with colour-related words whilst the aim of Experiment 2 was the same as Experiment 1. Although Experiment 1 showed no effects of repeating study-test sessions, it is still uncertain that no effect of repetition also occurs with colour-related words. Therefore, the present experiment still employed repeated study-test sessions. To achieve the aim, Experiment 2 used the colour-related words instead of colour names whilst the entire procedure was the same as the previous experiment. The hypothesis of Experiment 2 was the identical to Experiment 1; the highest memory test scores in the congruent condition and the lowest in the incongruent condition whilst the magnitude of the effects might be reduced as the number of study-test sessions increases.

Method

Participants

New 30 students of University of Kent at Canterbury were recruited as the participants. Five male participants (*M* = 18.80 years old; *SD* = 0.45) and 25 female participants (*M* = 20.00 years old; *SD* = 3.06) were recruited for the present experiment. These 30 students consisted of 24 native English speakers, five non-native English speakers, and one “English plus other language(s)” speaker.

Design

The research design was the same as Experiment 1 except that the participants studied pairs of colour-related words and numbers.

Materials

The only difference between Experiment 1 and 2 was the use of different colour words; the present experiment used eight colour-related words. Four colour-related words were obtained from the two original studies (Hazan-Liran & Miller, 2017; P. Miller et al., 2018) whilst new four words (i.e. *grape*, *lemon*, *carrot* and *cream*) were employed in the present experiment. Since the ink colour black was replaced with purple in the previous experiment, this experiment also used a purple-related word as a stimulus. The selection of purple-related word was based on Tanaka and Presnell’s (1999) study. In their first experiment the result revealed that 60 percent of 30 participants agreed purple as a typical image colour of the word *grape*. Therefore, the present experiment employed the word *grape* as the colour-related word for purple ink. Hazan-Liran and colleagues (Hazan-Liran & Miller, 2017; P. Miller et al., 2018) used the word *banana* as the yellow-related word. However, because the corpus (Johansson & Hofland, 1989) that was used in this experiment did not include the word *banana*, this word was replaced with *lemon* since Tanaka' and Presnell's first experiment showed 100 percent agreement for the *lemon*-yellow association. Additionally, this word length is five letters that is the average word length of the previous experiment in the present thesis. Therefore, the word *lemon* was used in the present experiment. In addition, the word *carrot* was used in Experiment 2 instead of the word *orange* (which used in the two original studies; Hazan-Liran & Miller, 2017; P. Miller et al., 2018). This is because the word *orange* might refer to a colour name (i.e. *orange*) rather than the fruit name (i.e. *orange*). In Tanaka and Presnell's (1999) experiment 100 percent of the participants agreed that orange is the typical colour for the word *carrot*. Thus, the word *orange* was replaced with *carrot* in the present experiment. Regarding the use of the white-related word, the two original research (Hazan-Liran & Miller, 2017; P. Miller et al., 2018) employed the word *whipped cream*. To balance the word length, the present experiment used the word *cream* as the white-related word. In terms of word frequency, since the corpus that was used in Experiment 1 did not include the selected eight colour-related words, the new corpus was used for the present experiment to reveal word frequencies. The word frequency in a million-word collection of present-day British English texts was obtained from *Frequency analysis of English vocabulary and grammar: based on the LOB Corpus / Vol. 1, Tag frequencies and word frequencies* (Johansson & Hofland, 1989). The selected eight colour-related words are listed here in a form of colour-related word (corresponding typical colour, the total number of occurrences in the corpus as a common noun): *sky* (blue, 53), *mud* (brown, 21), *grass* (green, 63), *carrot* (orange, 3), *strawberry* (red, 1), *cream* (white, 18) and *lemon* (yellow, 36). For the word *grape* (as the purple-related word), the total number of occurrences in the corpus is 2 as a common plural (*grapes*). The average frequency of these eight words was approximately 26.63 in the corpus and the raw average length of letters was 5.25 letters. These words were written in the same font type, ink colour, letter size and position with the same background screen as in Experiment 1. Differing from Experiment 1, new word-number pairs were created for the present experiment via the website (Randomness and Integrity Services Ltd, n.d.). The same word-number pairs were used across all participants (see Appendix B). The total number of study and test trials were the same as Experiment 1 and the same tone tasks were used as in the previous experiment.

Procedure

The procedure was the same as Experiment 1 except that the participants studied the colour-related words instead of the colour names. Experiment 2 also took approximately 15 minutes to complete.

Results

None of the participants were timed out. Four participants made responses in less than 200 ms; however, each of them made such responses only in one trial out of 72 test trials. Additionally, all participants' mean reaction times in each session of all colour congruency conditions were also longer than 200 ms. Since these could suggest the participants' appropriate attention to the task, all responses were included in the following analyses.

Analysis of Accuracy

Table 2.2 and Figure 2.2 showed a summary of descriptive statistics. As Experiment 1, the 2 (Colour Congruency) x 3 (Study-Test Session) factorial ANOVA was performed. This revealed that the main effect of Colour congruency [$F(1, 29) = 0.86$, $p = .363$, partial-$\eta^2 = .03$] and the interaction effect of Colour congruency x Study-test session [$F(1.508, 43.723)$

= 0.01, p = .991, partial-η^2 = .00] were not significant whilst the main effect of Study-test session reached at the significant level [F(1.480, 42.929) = 49.40, p < .001, partial-η^2 = .63].

Whilst the previous analysis reported no ink colour difference, the difference from the neutral condition might have differed between the congruent and incongruent conditions. Therefore, the baseline condition was included. The same 3 (Colour Congruency) x 3 (Study-Test Session) factorial ANOVA was conducted as in Experiment 1.

Table 2.2

Mean Score Proportions and Reaction Times (in ms) in Each Colour Congruency Condition (Standard Deviation in Parentheses; Experiment 2)

Condition	Study-Test Session	Accuracy	Reaction Time
		M (*SD*)	*M* (*SD*)
Congruent	1st	.46 (.18)	1,962 (474)
	2nd	.62 (.25)	2,101 (643)
	3rd	.77 (.21)	1,941 (882)
Incongruent	1st	.42 (.17)	2,074 (478)
	2nd	.58 (.22)	2,071 (591)
	3rd	.73 (.29)	1,853 (494)
Neutral	1st	.42 (.24)	2,030 (458)
	2nd	.68 (.24)	2,122 (714)
	3rd	.82 (.23)	1,861 (579)

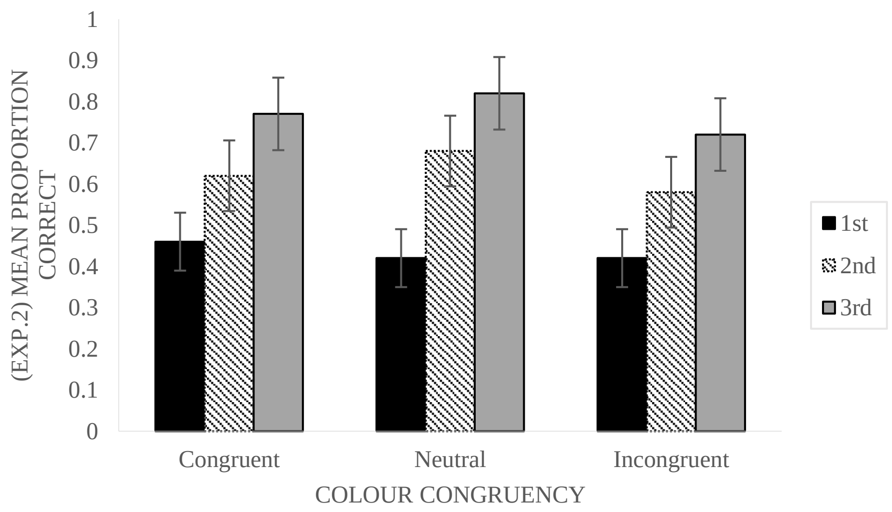


Figure 2.2. The graph shows the mean proportions of correct scores in each colour congruency condition (Experiment 2). Error bars represent the 95% confidence intervals (Masson & Loftus, 2003).

The results indicated that there was no main effect of Colour congruency, $F(2, 58) = 1.00$, $p = .375$, partial-$\eta^2 = .03$ [Congruent: $M = .61$, $SD = .17$; Incongruent: $M = .57$, $SD = .19$; Neutral: $M = .64$, $SD = .19$], and no interaction with Study-test session [$F(4, 116) = 1.09$, $p = .367$, partial-$\eta^2 = .04$]. As expected, the main effect of Study-test session was significant [$F(1.383, 40.110) = 73.40$, $p < .001$, partial-$\eta^2 = .72$] and paired sample *t*-tests (two-tailed) revealed that there was a significant increment from the first session to the second [$t(29) = 6.16$, $p < .001$, $d = 1.12$] and from the second to the third [$t(29) = 8.92$, $p < .001$, $d = 1.63$].

As Experiment 1, the pattern of incongruent ink colour combination as a between-participant factor was examined in the 3 (Colour Congruency) x 3 (Study-Test Session) factorial ANOVA; however, again the interaction effects with the incongruent ink colour combination pattern factor were not significant (all $p > .05$). Additionally, there was no memory accuracy difference between these seven patterns [$F(6, 23) = 1.29$, $p = .300$, partial-$\eta^2 = .25$]. These indicated that seeing the word-number pair with brown ink and yellow ink (as incongruent ink colours) had no impact on the magnitude of ink colour congruency.

For the effects from the previous trials (e.g. seeing the word *banana* established an incorrect link with yellow ink used in different pairs), since the previous experiment demonstrated that such effects had no impact on inducing ink colour effects with the AMST, these were not analysed in the present experiment. For the same reason, the patterns of incorrect responses (e.g. for the word *grape*, incorrectly typing the number *4* since this number had been presented in purple ink) were not analysed.

Discussion

Overall, Experiment 2 found that when the participants were asked to remember pairs of colour-related words and numbers with manipulation of ink colours printed in the numbers (e.g. *sky-1* with the number *1* in blue ink, brown ink or black ink), they could remember these pairs accurately regardless of ink colours whilst their memory accuracy scores increased as they studied and being tested more frequently. Interestingly, whilst the original studies (Hazan-Liran & Miller, 2017; P. Miller et al., 2018) found mismatching ink colours printed in the numbers produced interference effects, the main point of Experiment 2 is that using colour-related words on the AMST did not have such negative effects of incongruent ink colours on cued-recall memory. To examine whether the findings of Experiment 1 and 2 were restricted to a specific memory test type (i.e. cued-recall memory tests), it is necessary to explore the effects of task-irrelevant ink colours on the AMST with another type of memory tests; associative recognition tests.

Experiment 3

This experiment examined whether associative recognition tests can measure the effects of task-irrelevant ink colours on the AMST when the words on this task are colour names. Whilst Chapter 1 explained that the cued-recall tests and associative recognition tests are expected to reflect performance of paired-associate learning (Arndt, 2012), Experiment 1 used the cued-recall memory tests and revealed effects. Therefore, it is important to confirm that the ink colour effects on the AMST is not limited to cued-recall memory.

Bird (2017) explained that recognition memory can be defined as the ability to notice the previously-encountered information. The distinctive feature of recognition memory is that people do not have to recall specific information necessarily; it is the ability that people can remember “I have seen this before.” By introducing the dual-process signal detection model (DPSD; Yonelinas, 1994), Yonelinas, Aly, Wang and Koen (2010) noted that increasing the level of familiarity (i.e. the feeling of “I have seen this before”) of studied items can lead learners to correctly select these items on a recognition test. Therefore, recognition memory involves a recollection (i.e. remembering specific information of stimulus) factor and a familiarity aspect (Yonelinas et al., 2010). There are three types of recognition memory tests; a yes/no (or old/new) task, rating task and forced choice task (Macmillan & Creelman, 2005; Stanislaw & Todorov, 1999). The literature (Macmillan & Creelman, 2005; Stanislaw & Todorov, 1999) described that, in a yes/no (or old/new) task, after participants study items, they need to judge whether they have previously encountered the presented stimulus (“yes” or “old”) or the presented item was not in the studied list (“no” or “new”). In a rating task participants are asked to give a rating score to the presented stimulus whilst a forced choice task requires participants to select the studied item from the alternatives. When measuring the quality of paired-associate learning on a recognition task, Arndt (2012) noted that the typical research design is an associative recognition task; after participants study item pairs (e.g.

stove-letter), they are asked to distinguish the correct item pairs from the incorrect item pairings (e.g. *stove-dance*). Note that, in this task, the words composed of the incorrect pairs are both presented previously; just items are incorrectly paired. Since following this design with yes/no (or old/new) judgment would be the simplest to measure recognition memory in the case of the present research, an associative recognition task with correct/incorrect judgment was employed for this study.

When considering how to analyse recognition memory performance, applying signal detection theory (for a review see Stanislaw & Todorov, 1999) would allow the present research to measure such memory performance precisely, especially when with yes/no (or old/new) judgment. Whilst the simplest way to analyse recognition memory performance would be analysing the hit rate (i.e. the probability of answering "yes" or "old" to the studied items) and the false alarm rate (i.e. the probability of judging "yes" or "old" to the unstudied items), it is difficult to precisely analyse memory accuracy whilst removing the influence of response bias (i.e. the tendency to answer "yes" or "no" judgment) in this way of analyses. To do this, a measurement d' (d-prime) values can be used. Since d' values can reflect recognition memory accuracy despite participants' individual preferences for answering old/new (Ingham, 1970), this index was employed in the present research. As the literature (Stanislaw & Todorov, 1999) described, a greater d' value indicates more superior recognition memory accuracy. Analysing response bias is also important since it becomes possible to determine whether the high hit rate (i.e. correctly judging "old" to old items) is due to the high ability of distinguishing old items from new stimuli or the high tendency to answer "old" (Ingham, 1970; Stanislaw & Todorov, 1999). The present research employed the index c to represent participants' response bias since c values are not affected by changes of d' values (for a summary see Stanislaw & Todorov, 1999).

In the present experiment performance on the recognition tests was expected to reflect the effects of colour congruency on the AMST based on the study of internally activated working memory contents conducted by Kiyonaga and Egner (2014), which research design appeared to be close to the AMST. Internal activation refers to processing information that is formerly existed and now disappears from the environment. To test the effects of internal activation of working memory, they developed a new paradigm, which was named the working memory Stroop task. In this task the researchers first presented the colour name (e.g. *blue*) in black ink and instructed participants to remember the word for a later memory test. After this word disappeared, the participants were asked to name the ink colour of the rectangular patch, which were either congruent (i.e. blue ink) or incongruent (e.g. green ink) to the previously-encountered colour name. Upon completion of this, the colour name was presented and the participants were asked to judge whether they had seen the presented colour name or not. In this way, the researchers examined how colour congruency between the internally activated colour image of “blue” by seeing the word *blue* and the externally activated colour image of “green” by seeing green ink affects performance on the colour-naming task and the recognition memory test. The results demonstrated that the participants were less accurate on the recognition memory tests in the incongruent trials than in the congruent trials, which confirmed that semantic congruency between internal and external affects one’s recognition memory accuracy. The researchers explained that since internally attending to information in working memory and externally doing it in the environment requires the same cognitive resources, filtering incongruent information in the incongruent trials used the resources and the less amount of these was left for maintaining the working memory contents. Since the research design of the present research (i.e. internally activating colour concepts by reading colour words printed in black ink whilst seeing ink colours physically by seeing numbers and memory tests are conducted later) appears to be close to

Kiyonaga and Egner's (2014) research design, if their resource sharing hypothesis could be applied to the present experiment, seeing and filtering out the task-irrelevant incongruent ink colours (e.g. *blue-5* with the number *5* in brown ink) was expected to divert the cognitive resources from remembering the word *blue* (and/or the number *5*). Therefore, recognition tests are also assumed to be capable to reflect the effects of task-irrelevant ink colours on the AMST.

Therefore, Experiment 3 examined the question; whether it is possible to observe the identical ink colour effects as the two original studies (Hazan-Liran & Miller, 2017; P. Miller et al., 2018) and Experiment 1 in the present thesis when the AMST is combined with an associative recognition task and the words on the AMST are colour names. The current experiment aimed to test whether a task-irrelevant ink colour printed in a number and its association with a colour name in the AMST affect accuracy of recognition memory for word-number pairs when the analyses are based on d' values and c values. To achieve this aim, the participants were asked to remember eight pairs of colour names and numbers (e.g. *blue-7*) as in Experiment 1 in the present thesis and later these memories were tested on associative recognition memory tests, which asked the participants to judge whether they had studied the presented word-number pairs before or not. Differing from Experiment 1 and 2, the participants experienced one study-test session in each of three colour congruency conditions (i.e. the congruent, incongruent and neutral conditions). This was since the interaction effects of colour congruency and study-test sessions in Experiment 1 and 2 were not significant whilst the colour names revealed effects in Experiment 1; there would be no reason to examine the practice effect with colour names. The numbers of correct and incorrect responses were recorded for the analyses. Since Experiment 1 revealed that the magnitude of interference was more robust than of facilitation and the literature (Kiyonaga & Egner, 2014) suggested the underlying processes for interference, Experiment 3 expected that interference would be more likely to occur than facilitation; the d' value would be lower in the incongruent condition compared to the neutral condition.

Method

Participants

Participants were new 30 students at University of Kent at Canterbury. Two male participants (*M* = 19.00 years old; *SD* = 1.41) and 28 female participants (*M* = 19.32 years old; *SD* = 1.34) were recruited for the present experiment. Thirty participants consisted of 16 native English speakers, 12 non-native English speakers and two "English plus other language(s)" speakers.

Design

Experiment 3 examined ink colour effects on the AMST via an associative recognition, which asked participants to judge whether the presented pairs were the same as the studied pairs or different. The current experiment was a within-participant design with Colour congruency (Congruent, Incongruent and Neutral) as a within-participant factor. The other parts of the design were identical to Experiment 1 and 2.

Materials

The majority of the materials in Experiment 3 were obtained from Experiment 1. From the two original studies (Hazan-Liran & Miller, 2017; P. Miller et al., 2018), the first four practice word-number pairs were employed for the practice section (see Appendix C). These practice words and numbers were presented in black ink. The order of all word-number pairs across the experiment was randomised by Psychopy (Peirce et al., 2019) for each participant. Regarding the combination pattern of colour names and incongruent ink colours, the present experiment employed only one combination pattern (e.g. all participants saw the *blue-5* pair with the number *5* was written in brown ink; see Appendix D). This was because

Experiment 1 reported that the pattern of incongruent colour-word combination had little impact on the interaction effect of Colour congruency and Study-test session. Additionally, the two original studies (Hazan-Liran & Miller, 2017; P. Miller et al., 2018) seemed to use only one combination pattern and revealed ink colour effects. Therefore, the present experiment was conducted with only one combination pattern in order to make the present research design closer to the original methodology.

New pairs of colour names and numbers were created as studied pairs with random pairings (see Appendix D). The reason why the present experiment employed different to-be-remembered word-number pairs was to examine whether the observed ink colour effect in Experiment 1 is limited to specific word-number pairings or it can happen with any word-number pairings. The incorrect (i.e. distractor) word-number pairs were different combinations of words and numbers (e.g. when the participants studied the *blue-5* pair in the study section, they saw the *blue-4* pair as the incorrect pair in the test section). To create these incorrect pairs, each colour name was randomly paired with one of the other seven numbers (since one number had been already used to create the studied pairs). The fixed word-number pairs were used across all participants.

Procedure

Whilst the basic procedure of Experiment 3 was identical to Experiment 1 and 2, there were several changes made in the general instruction, the practice section and the test section. In the beginning of Experiment 3, the instruction on the screen asked the participants to use the “C” key with the left hand’s forefinger for the correct (i.e. studied) pairs and the “I” key with the right hand’s forefinger for incorrect (i.e. distractor) pairs on a later recognition test. After the instruction, the participants experienced eight practice trials with the feedback in the practice section (see Appendix C). The four practice words were neutral words (e.g. *song*), which were paired with numbers (e.g. *8*). These were presented in black ink with the same exposure duration, size and location as the studied pairs. The participants saw these four word-number pairs one at a time. Upon the completion of seeing all four practice pairs, the participants were asked to make correct/incorrect responses for the presented word-number pairs (which were also presented one by one). These practice word-number pairs were also kept visible until the participants made responses. When the participants made a correct response, the feedback "Correct!" was shown on screen for 1,000 ms whilst the feedback for the incorrect key response was "Oops! That was wrong." During these eight practice trials, half pairs required the correct responses and half were the incorrect pairs. When the participants finished this practice section, the screen explained that they would experience the study section, which would be followed by 16-trial (8 correct + 8 incorrect pairs) recognition test and 2 minutes would be given for this test section. In the test section the task was to judge whether they had seen the presented word-number pairs before or not with the assigned keys (i.e. the "C" key or the "I" key). The words and numbers were presented in black ink and one at a time in the test section. This section automatically ended for completion of all 16 trials or out of time. In total, there were 48 test trials [(8 correct pairs + 8 incorrect pairs) x 3 colour congruency conditions] plus the two 16-trial tone comparison tasks. The present experiment took approximately 10 minutes to complete.

Results

Mean hit (Hit) rates, false alarm (FA) rates, d' values, c values and reaction times in each colour congruency condition are shown in Table 2.3 and Figure 2.3.

Table 2.3

Hit Rates, False Alarm Rates, d' Values, c Values and Reaction Times (in ms) in Each Colour Congruency Condition (Experiment 3)

Condition	Hit Rate	FA Rate	d'	c	Reaction Time
	M (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)
Congruent	.69 (.18)	.36 (.19)	1.01 (1.05)	-0.06 (0.27)	1,677 (409)

Incongruent	.57 (.17)	.44 (.17)	0.35 (0.71)	-0.01 (0.34)	1,719 (439)
Neutral	.69 (.15)	.35 (.18)	0.96 (0.84)	-0.06 (0.26)	1,618 (392)

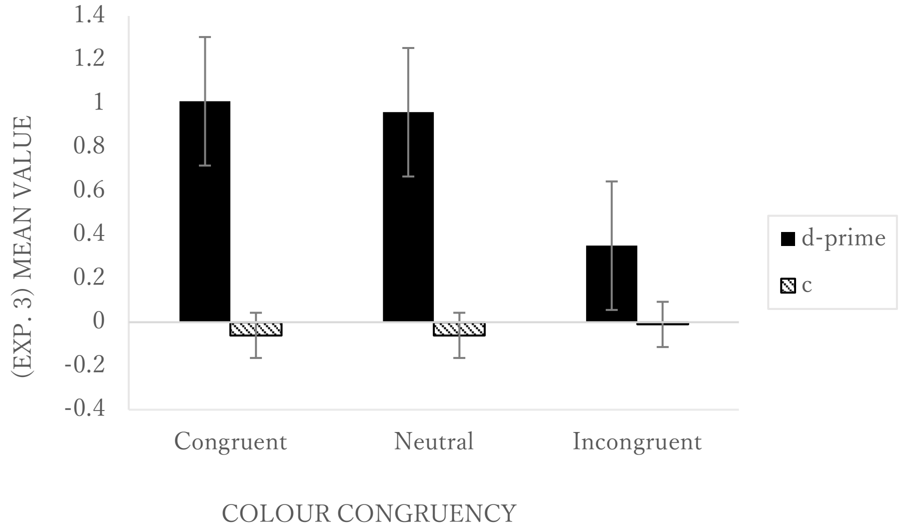


Figure 2.3. The graph shows the mean d' values and the mean c values in each colour congruency condition (Experiment 3). Error bars represent the 95% confidence intervals (Masson & Loftus, 2003).

None of the participants were timed out in the test section. Two participants made responses in less than 200 ms; however, each of them made such responses only in one trial out of 48 test trials. Additionally, all participants' mean reaction times in each colour congruency condition were also longer than 200 ms. Since these could indicate that all of them paid appropriate attention to the task, all responses were included in the following analyses.

Analysis of Accuracy

A one-way ANOVA with Colour congruency (Congruent, Incongruent and Neutral) as a within-participant factor was conducted. d' values were calculated following the formulae in the literature (Macmillan & Creelman, 1991) whilst the Hit rates were calculated by Hit rate = (the number of Hit+.5) / (the number of Hit + the number of Miss + 1) and the FA rates were calculated by FA rate = (the number of FA+.5) / (the number of FA + the number of CR + 1) (Joan Gay Snodgrass & Corwin, 1988). Based on these rates, d' values and c values were calculated (see Table 2.3 and Figure 2.3).

Analysis of d' values indicated that there was a main effect of Colour congruency, $F(2, 58) = 6.25$, $p = .003$, partial-$\eta^2 = .18$. Paired sample *t*-tests (two-tailed) indicated that the incongruent condition scored significantly lower than the congruent condition [$t(29) = 3.32$, $p = .002$, $d = 0.61$] and the neutral condition [$t(29) = 3.09$, $p = .004$, $d = 0.56$]. Memory accuracy in the congruent condition and the neutral condition was identical, $t(29) = 0.21$, $p = .832$, $d = 0.04$. The results indicated that incongruent ink colours made it difficult to differentiate the studied and unstudied word-number pairs.

The same one-way ANOVA with response bias measure (c) revealed no significant main effect of Colour congruency [$F(2, 58) = 0.32$, $p = .724$, partial-$\eta^2 = .01$], which indicate the reported results of accuracy was not affected by response bias. Although all colour congruency conditions showed negative response bias values (i.e. response tendency to answer "old"; for a review see Stanislaw & Todorov, 1999), this could be because the unstudied (i.e. distractor) pairs were just different pairings of previously seen items.

Discussion

Overall, Experiment 3 found that when the participants studied pairs of colour names and numbers with the incongruent ink colours (e.g. the *blue-5* pair with the number *5* in brown ink), they showed the poorest recognition memory accuracy (i.e. the ability to correctly answer "correct" to the studied word-number pairs and "incorrect" to the unstudied word-number pairs) amongst three colour congruency conditions (i.e. the congruent, incongruent and neutral conditions). However, it is still uncertain whether this ink colour effect can be transferred when the AMST uses a different word type; colour-related words. Therefore, Experiment 4 explored this issue; whether it is possible to observe the identical ink colour effect when the words on the AMST are colour-related words.

Experiment 4

Experiment 4 explored whether the ink colour effect is observed with associative recognition tests when the words on the AMST are colour-related words. The aim of the present experiment was to test whether a task-irrelevant ink colour printed in a number and its relationship with a colour-related word affect recognition memory accuracy of word-number pairs. Since Experiment 1 and 3 reported the interference effect at a significant level ($p < .05$), it was expected that this type of ink colour effects might be more likely to be observed than the facilitation effect if any effects happen. Therefore, the hypothesis was that the incongruent condition would reveal poorer recognition memory accuracy than the neutral condition.

Method

Participants

Participants were new 30 students at University of Kent. Two male participants (*M* = 18.50 years old; *SD* = 0.71) and 28 female participants (*M* = 20.18 years old; *SD* = 4.55) were recruited for the present experiment. The 30 participants consisted of 16 native English speakers, 11 non-native English speakers and three “English plus other language(s)” speakers.

Design

The research design was the same as Experiment 3 except that the participants studied the colour-related words in the present experiment.

Materials

The majority of the materials was obtained from Experiment 2 (see Appendix B). Since the present experiment attempted to replicate the methodology of Experiment 2 with a different memory test (i.e. recognition test), it was ideal to use the same materials so the only difference between Experiment 2 and the present experiment was a type of memory test. By doing this, it was possible to examine whether using the cued-recall test could explain why ink colour effects were not observed in Experiment 2. Therefore, the present experiment used the same word-number pairs for correct pairs as used in Experiment 2. Regarding the combination pattern of words and incongruent ink colours, the present experiment used all seven combination patterns (e.g. Participant 1 saw the *blue-1* pair with the number *1* was written in yellow ink, Participant 2 saw the same pair but *1* was in white ink and so forth). This was due to the wider variety of congruent ink colours with colour-related words compared to colour names. For example, presenting the *blue-5* pair with brown ink could be a complete incongruent condition. However, presenting the *sky-5* pair with white ink might induce a picture of the blue sky with white clouds, which would not be a complete incongruent condition. Therefore, presenting the same word-number pairs with various incongruent ink colours was necessary to reduce this risk in the present experiment.

The incorrect pairs were created by randomly pairing of colour-related words and numbers (see Appendix B). The same word-number pairs were used across all participants (e.g. all participants studied the *sky-1* pair in the incongruent condition).

Procedure

The procedure was the same as Experiment 3 except that the participants were asked to study pairs of colour-related words and numbers. The present experiment also took approximately 10 minutes to complete.

Results

None of the participants were timed out in the test section. Two participants made responses in less than 200 ms; the one who made such responses only in two trials out of 48 test trials and the other participant who made such responses in two trials of the neutral condition and in seven trials of the incongruent condition. Although the latter person showed the high number of extremely quick responses, none of the reported conclusions were modified by excluding this one person. Therefore, all participants' responses were included in the following analyses. Mean hit (Hit) rates, false alarm (FA) rates, d' values, c values and reaction times in each colour congruency condition are summarised in Table 4 (also see Table 2.4 and Figure 2.4).

Table 2.4

Hit Rates, False Alarm Rates, d' Values, c Values and Reaction Times (in ms) in Each Colour Congruency Condition (Experiment 4)

Condition	Hit Rate	FA Rate	d'	c	Reaction Time
	M (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)
Congruent	.67 (.13)	.36 (.16)	0.88 (0.80)	-0.04 (0.22)	1,785 (510)
Incongruent	.65 (.20)	.36 (.18)	0.87 (1.10)	-0.01 (0.30)	1,668 (605)
Neutral	.73 (.17)	.35 (.19)	1.15 (1.09)	-0.13 (0.22)	1,615 (560)

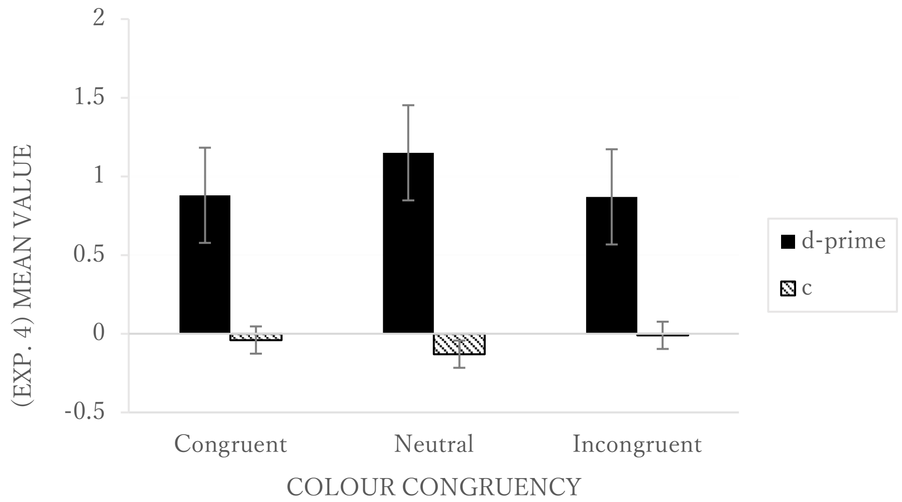


Figure 2.4. The graph shows the mean d' values and the mean c values in each colour congruency condition (Experiment 4). Error bars represent the 95% confidence intervals (Masson & Loftus, 2003).

Analysis of Accuracy

The same one-way ANOVA as Experiment 3 was conducted. The analysis of d' values indicated that there was no significant main effect of Colour congruency, $F(2, 58) = 1.10$, $p = .339$, partial-$\eta^2 = .04$. Even when the congruent and incongruent conditions were compared, paired sample *t*-test (two-tailed) showed no significant difference, $t(29) = 0.02$, *p*

= 982, *d* = 0.004. This could suggest that when the participants studied the pairs of colour-related words and numbers (e.g. *sky-4*), the ink colours of the numbers during studying had no impact on their recognition memory accuracy for word-number pairs.

The same one-way ANOVA with c values (which represent response bias) revealed no significant main effect of Colour congruency, $F(2, 58) = 2.07$, $p = .135$, partial-$\eta^2 = .07$, whilst the analysis revealed negative c values in all colour congruency conditions as Experiment 3.

Discussion

The present experiment replicated the outcomes as in Experiment 2; the participants remembered the pairs of colour-related words and numbers at the equal levels of accuracy regardless of task-irrelevant ink colours printed in the numbers during study.

Chapter 2: Chapter Summary

The present chapter tested the applicability of the AMST to examine memory performance. There were four research questions that the present chapter explored; (1) whether the findings of the original studies (Hazan-Liran & Miller, 2017; P. Miller et al., 2018) can be extended to memory accuracy, (2) if the effects are observed, which direction the task-irrelevant ink colours affect memory (i.e. facilitation and/or interference), (3) whether the effects occur with both colour names and colour-related words and (4) whether the type of memory tests (i.e. cued-recall or associative recognition) modifies the robustness of the effects. The aim of this chapter was to examine whether the effects of task-irrelevant ink colours and their relationships with colour words on the AMST produce memory enhancement or memory impairment when these effects are tested by the cued-recall tests and the associative recognition tests. The overall hypothesis was that the congruent condition would score higher than the neutral condition (i.e. facilitation) whilst the incongruent condition would score lower than the neutral condition (i.e. interference) on both memory tests. To test the hypotheses, the participants were asked to remember the presented pairs of colour names (Experiment 1 and 3) or colour-related words (Experiment 2 and 4) and numbers whilst only the numbers were printed in colours (e.g. red ink). After that, they were instructed to type the paired number when the colour word appeared on screen one at a time (i.e. a cued-recall test; Experiment 1 and 2) or to judge whether they had seen the presented word-number pair or not (i.e. an associative recognition test; Experiment 3 and 4). During these memory tests, all stimuli were presented in black ink. Data from these four experiments revealed the following results; (1) interference was observed with the colour names on both memory tests, (2) facilitation was significant with the colour names on the cued-recall tests when analysing performance in the second and third study-test sessions, (3) the robustness of these effects were not interacting with the number of study-test sessions and (4) no ink colour effects were observed with the colour-related words on both memory tests. Therefore, the hypotheses were partially supported by the results with the colour names whilst the experiments with the colour-related words did not support it. These findings confirmed that the effects reported in Hazan-Liran and colleagues' (Hazan-Liran & Miller, 2017; P. Miller et al., 2018) studies occurred due to difference in learning efficacy, not due to visual search strategies.

The reported results with the colour names are partially consistent with the findings of the two original studies (Hazan-Liran & Miller, 2017; P. Miller et al., 2018) whilst with the colour-related words are different. The first result could be explained in the framework of cognitive load theory as Hazan-Liran and colleagues described. That is, seeing incongruent ink colours automatically activated colour concepts (even though it was irrelevant to the task). Since the suppression of this task-irrelevant information used up cognitive resources, less available cognitive resources were left for learning the association between the colour names and numbers; thus, memory accuracy in the incongruent condition was degraded.

For the facilitation effects, since seeing congruent ink colours triggered colour concepts that were related to the to-be-remembered colour names, there was no need to suppress this information although it was task irrelevant. Rather, activation of the same colour concepts would assist learners to produce the stronger memory traces of the colour names. Thus, memory performance was the most superior in the congruent condition. Since this effect was observed in the second and third sessions, it could indicate that occurrence of facilitation may require some experience of the task. However, the second result raises a question; why facilitation was not observed on the associative recognition tests. The possible explanation would be since memory accuracy in the neutral condition became more superior on this type of memory test compared to the cued-recall tests according to encoding specificity principle (Tulving & Thomson, 1973). This principle assumes that retrieval is enhanced when features of information at encoding match features of contexts at recall. In Experiment 3 the participants in the neutral condition saw the word-number pairs in black ink at the study section and at the test section. Since only the neutral condition presented the learning materials with the same features (i.e. black ink) at encoding and retrieval, it might have assisted learners' retrieval processes. When considering this potential scenario, one might question whether significant difference between the neutral and incongruent conditions on the recognition tests occurred due to increased accuracy in the neutral condition rather than due to degraded learning efficacy with incongruent ink colours.

Answering this question would be difficult since what can be an appropriate baseline condition is a controversial issue in the Stroop research. For instance, the literature (Salo, Henik, & Robertson, 2001) demonstrated how participants' responses were affected by different type of the neutral Stroop stimuli. Salo et al. employed three types of neutral stimuli on the traditional Stroop task: nonword-physically-similar strings (e.g. *XXX*), nonword-physically-dissimilar strings (e.g. *XXX*, *WWWW*, *SSSS*) and animal names (e.g. *dog*, *tiger*). The analyses revealed the fastest response to the nonword-physically-similar strings, the second fastest to the nonword-physically-dissimilar strings and the slowest to the animal names. This illustrated that features of neutral stimuli might modify differences between coloured conditions and a control condition. Therefore, if different underlying processes are assumed to contribute to facilitation and interference, there would be some manipulation that affects only one of the two effects. However, in the present research, it is noteworthy to highlight that interference was also observed on the cued-recall tests, which overlapping the ink colours of the neutral stimuli at study and test did not occur. That is, the present research would favour the assertion that ink colour congruency effects would affect when learners study items rather than when they recall information. Thus, whilst what manipulation (e.g. what can be ideal neutral stimuli) modifies the robustness of facilitation or interference is open for future research to explore, the current thesis kept using the pairs of colour concepts and numbers printed in a neutral ink colour as a baseline condition.

The third result showed contradicting evidence against the literature (Wixted, 2005), which signalled that the strength of task-irrelevant ink colours might change with practice. However, the stable ink colour effects across practice might suggest that ink colours produce effects more on encoding than on retrieval. The possible account might be that in the present research the participants had several study-test sessions to train binding the words and numbers. However, the instruction did not mention about suppression of task-irrelevant ink colours. Since the participants might have questioned about what to do with ink colours, they may have keeping paying attention to the ink colours throughout the experiments. Thus, practice may have not produced significant effects on the magnitude of colour congruency in the present research. As the goal of the present research is to examine whether the AMST can be a paradigm to affect the quality of paired-associate learning, dissociation of ink colour effects at encoding and at retrieval may be something that future research can explore.

Contradicting evidence from the present chapter against the two original studies (Hazan-Liran & Miller, 2017; P. Miller et al., 2018) hampers from determining whether colour-related words on the AMST produce effects on memory performance. Since the only difference between the experiments with the colour names and the colour-related words was semantic relatedness to associate colour concepts in the present research, it seems difficult to explain the reported results only from the cognitive load perspective. Rather, the present research would suggest to add semantic components to the existing model of cognitive load theory. When comparing effects between colour names and colour-related words with the Stroop task, one typical factor to consider would be the number of semantic neighbours (i.e. the concepts that are connected to the target concept; for a summary see Levin & Tzelgov, 2016). However, there are other factors that have revealed effects when studying memory.

For instance, in memory research, word concreteness (e.g. Paivio, 1965), the experience of sleep (e.g. Diekelmann & Born, 2010), previous encounter of stimuli (i.e. priming; e.g. Gagnepain, Lebreton, Desgranges, & Eustache, 2008) and the stimulus presentation format (e.g. Defeyter, Russo, & McPartlin, 2009) have been demonstrated effects on memory performance. Therefore, before making a conclusion that including semantic neighbours is recommended to the existing model of cognitive load theory, it will be necessary to examine which factor might be necessary to be included other than the number of semantic neighbours.

Additionally, an emotional factor might be a potential factor that the existing model of cognitive load theory needs to include. For example, Plass and Kalyuga (2019) argued that cognitive load theory needs to consider the effects of emotions with the idea that processing task-irrelevant emotional states create additional extraneous cognitive load. They also summarised the effects of emotions in memory research. In the present chapter such a factor might have contributed to memory impairment in the incongruent condition. This assumption was suggested based on the notion that incongruent ink colours in the Stroop-related tasks induce negative emotions (for a summary see Damen, Strick, Taris, & Aarts, 2018). That is, since the word-number pairs presented in the incongruent conditions induced negative emotions from the participants, they needed to suppress these task-irrelevant emotions that required additional cognitive resources; consequently, the less available cognitive resources were left for binding the two items. However, further study is required to confirm this hypothesis.

Data from the present chapter could imply that when semantics of learning materials and of to-be-ignored materials are robustly related to each other, to-be-ignored materials affect learning efficacy. Optimising accuracy of information acquisition would be achieved when considering semantic relationships between learning materials and these contextual features.

Importantly, this was the first study that applied the AMST to examine the effects of task-irrelevant ink colours and their associations with colour words on memory accuracy. The present study extended the findings of the two original studies (Hazan-Liran & Miller, 2017; P. Miller et al., 2018) by demonstrating no interference from incongruent ink colours with the colour-related words, which suggests a new possibility that undesirable task-irrelevant ink colour effects on learning can be neutralised when the semantic relationship between ink colours and colour words is indirect. However, it still remains unsolved what could be a factor that the existing model of cognitive load theory needs to include. This might be solved by exploring the features of colour-related words and influence of emotions that might have contributed to the effects of task-irrelevant ink colours on the AMST. Therefore, Chapter 3 examined the former factors whilst Chapter 4 tested the emotional factor with the AMST.

Overall, it was found that when studying pairs of colour names and numbers, memory tests could reflect effects from task-irrelevant ink colours whilst ink colour effects with colour-related words were not observed. Whilst the use of colour-relate words needs further investigation, the former part of the finding demonstrates a favour in the idea that the AMST can be considered as a valid and reliable paradigm to study the task-irrelevant ink colour effects on learning efficacy. Additionally, the reported findings seem to emphasise the consideration of semantic relationships between to-be-remembered information and ignored materials to optimise learners' ability of precisely acquiring information. Since evidence from the present experiments may suggest to conduct further study with colour-related words and to examine the contribution of emotions to memory performance, the next chapter conducted a series of experiments with colour-related words in order to test what would be a possible factor to be included in the existing model of cognitive load theory.

Chapter 3: Exploration of Colour-Related Words on the AMST with Recognition Tests

This chapter explores what could be a potential factor to be included in the model of cognitive load theory. Since the previous chapter demonstrated different results between colour names and colour-relate words, it was suggested that the contents of cognitive load theory may need to consider an additional factor to explain the effects of task-irrelevant ink colours on the AMST. The number of semantic neighbours (for a summary see Levin & Tzelgov, 2016) could be one potential factor from the perspective of the Stroop research. However, as Chapter 1 introduced, there seems to be other factors that have revealed effects on paired-associate learning and associative memory dimensions: word concreteness (e.g. Paivio, 1965), the experience of sleep (e.g. Fenn & Hambrick, 2012, 2013; Potkin & Bunney, 2012), previous encounter of stimuli (i.e. priming; e.g. Spieler & Balota, 1996) and the stimulus presentation format (e.g. Paivio & Yarmey, 1966). Therefore, the present chapter examined each factor with colour-related words since if any ink colour effect is observed, that factor might be able to explain why no ink colours were observed with colour-related words in the previous chapter; thus, it would be the one that the existing model of cognitive load theory needs to include.

The overall research question of Chapter 3 was whether there is a circumstance that memory tests can reflect effects from task-irrelevant ink colours and its relationship with colour-related words on the AMST. The present chapter aimed to study whether manipulation of word concreteness, the experience of sleep, previous encounter of stimuli and the stimulus presentation format can modify the magnitude of ink colour congruency with colour-related words, which effects are measured by the associative recognition tests. The present chapter employed this type of memory test since this test requires a fewer response keys (i.e. the “C” key and the “I” key) than the cued-recall tests. To achieve the aim, Experiment 5 tested the word concreteness account by using emotional words (Experiment 5-A was conducted to determine which emotional words were more likely to produce effects on memory performance and Experiment 5-B used one type of emotional words based on the previous experiment) whilst Experiment 6 studied the effects of sleep by creating the research design as a two-day experiment. Experiment 7 examined the effects of priming by comparing semantic priming and lexical priming whilst Experiment 8 presented colour concepts in the pictorial format and the verbal format to evaluate the effects of the stimulus presentation format. The overall hypothesis was that through these manipulations, colour-related words would become able to reveal ink colour effects when effects are measured by memory tests. Specifically, since interference was consistently observed in the previous experiments, it was assumed that interference (i.e. poorer memory accuracy in the incongruent condition than the neutral condition) would be more likely to occur compared to facilitation (i.e. higher memory accuracy scores in the congruent condition than the neutral condition).

Experiment 5-A

Experiment 5-A was conducted to determine what type of colour-related words on the AMST are more likely to produce effects on memory performance. In Chapter 1 no ink colour effects with colour-related words contradict the findings of the original studies (Hazan-Liran & Miller, 2017; P. Miller et al., 2018). One explanation for this would be colour-related words might require greater power to reveal effects since the effects of this type of words are smaller than of colour names. Or possibly, if no ink colour effects are true, the involvement of mental imagery might relate to such effects based on the literature of paired-associate learning introduced in Chapter 1 (e.g. Paivio, 1965). Specifically, constructing mental pictures with incongruent ink colours might modify the encoding processes with incongruent ink colours. This influence is more likely to occur with colour-related words than colour names due to the higher word concreteness of colour-related words. The literature has reported the superiority of concrete words over abstract words on cognitive tasks (i.e. The concreteness effect; Kousta, Vigliocco, Vinson, Andrews, & Del Campo, 2011; Paivio, 1991, 2006, 2013) since the higher ability of concrete words to induce mental imagery contributes to this advantage (Paivio, 1965, 1991, 2006; Paivio & Csapo, 1969; Paivio, Yuille, & Smythe, 1966; Yarmey & Paivio, 1965). In Chapter 1 when analysing the word concreteness values [using the ratings from Brysbaert, Warriner, and Kuperman’s (2014)] between the selected eight colour names and colour-related words, the colour-related words significantly showed the higher concreteness values than the colour names, $t(14) = 5.62$, $p < .001$, $d = 2.81$. Thus, in the case of seeing the *strawberry-7* pair with white ink in the incongruent condition, the participants could construct a mental image of strawberries on top of whipped cream despite incongruency between the colour-related word (i.e. *strawberry*) and the ink colour (i.e. white ink). Consequently, the robustness of ink colour incongruency could not be strong enough to produce effects, which resulted in no ink colour effects with colour-related words.

This waning was less likely to happen in the original studies (Hazan-Liran & Miller, 2017; P. Miller et al., 2018). The beneficial effects of concrete words appear to require enough time for producing mental imagery (Paivio, 1991; Paivio & Csapo, 1969). However, in Hazan-Liran and colleagues' studies, the participants were asked to respond as fast as possible and the studied items in the table were available at the top of the page during the task, the participants would not have any reason to rely on mental imagery. Therefore, it is still questionable whether mental imagery affects performance on the AMST when memory tests are used to examine the effects of task-irrelevant information on the AMST.

Although it intuitively seems to be difficult to find "abstract" colour-related words since colour-related words generally refer to tangible and/or observable concepts, emotional words can be abstract colour-related words according to the literature. Evidence for the relationship between colours and emotion has been reported by Valdez and Mehrabian (1994). Their study aimed to explore how people's emotional reaction might be different for three colour manipulations: hue, saturation and brightness. According to Valdez and Mehrabian's definition, hue refers to wavelength. This represents what colour names refer to. The examples of short-wavelength hues are the colour "blue" and the colour "green" whilst long-wavelength hues are the colour "yellow" and "orange." Saturation (i.e. chroma) is the level of vividness or greyness. For instance, lower saturation means that the colour contains more grey. Brightness represents the level of black-to-white. The researchers manipulated these three components of colours and recorded participants' emotional reactions using rating scales based on The Pleasure-Arousal-Dominance (PAD) Emotion Model (Mehrabian & Russell, 1974, as cited in Valdez & Mehrabian, 1994), which assumes that emotion can be divided into three dimensions: pleasantness, arousal and dominance. In their study the participants saw colour samples one at a time and were asked to give rating scores to the presented colour samples on pleasure, arousal and dominance scales. Interestingly, the results showed a congruent pattern between three aspects of colours and emotional reactions, such as the participants saw the colour “blue” as the most pleasant colour, arousal increased as the level of saturation became higher and more saturated and less bright colours produced the higher level of dominance. Therefore, Valdez and Mehrabian’s (1994) use of the PAD model provided the precise effects of colours on emotion beyond the simple emotion label such as “positive” or “dislike.”

In addition to Valdez and Mehrabian’ research, a further study was conducted by Sutton and Altarriba (2016). Sutton and Altarriba’s research aimed to create a list of emotional words and its associated colours. In their research they categorised emotional words into two types: emotion words and emotion-laden words. Emotion words are the direct emotion labels (e.g. *happy*, *sad*) whilst emotion-laden words are the words that are indirectly associated with certain emotional states (e.g. *death*, *pretty*). The task was that when the participants saw each emotional word, they were asked to write down the first colour that they could think of. Their results were summarised in the list, which showed two types of emotional words (i.e. emotion words and emotion-laden words), its associated colours and the frequency of each associated colour (e.g. for the word *Angry*, the colour “red” was most frequently reported as the first colour). The researchers found the clear patterns for the negative emotional words and positive emotional words; the colour “red” was the most common first-reported colour amongst both negative emotion words and negative emotional-laden words whilst “yellow” was the most common colour for the positive emotion words but “white” was the most frequently-reported first colour for the positive emotional-laden words. Having combined evidence from Valdez and Mehrabian (1994) and Sutton and Altarriba

(2016), it was confirmed that seeing colours can induce certain emotional reactions and reading emotional words can lead people to think about particular colours.

Whilst evidence from the literature (Sutton & Altarriba, 2016; Valdez & Mehrabian, 1994) seems to provide support for the idea that using emotional words as abstract colour-related words, the original studies conducted by Hazan-Liran and Miller (2017) and P. Miller, Hazan-Liran and Cohen (2018) employed only colour names and colour-related words. Therefore, it is still unknown whether using emotional words as abstract colour-related words in the AMST can produce task-irrelevant ink colour effects on encoding. To solve this issue, Experiment 5-A and 5-B examined whether increasing abstractness of colour words contributes to the production of task-irrelevant ink colour effects when such effects on the AMST are measured by memory tests.

Experiment 5-A explored the question of which type of emotional words (i.e. positive and negative) are appropriate when studying memory accuracy with the AMST. When exploring the use of emotional words with manipulating colour congruency, the one issue of concern is the complexity of the research design. Since there are three colour congruency conditions (i.e. congruent, incongruent and neutral) and emotional words can be categorised into three types (i.e. positive, negative and neutral), it would be a complex and long study, which might jeopardise task performance on memory tests due to fatigue whilst following the within-participant design in the original studies (Hazan-Liran & Miller, 2017; P. Miller et al., 2018). Thus, the aim of the present experiment was to determine which type of emotional words would show any effects on memory accuracy with the AMST before considering the interaction effect of colour congruency and emotion. To achieve this aim, the participants saw the pairs of colour names and “emotion” words (e.g. *happy*, *sad*) when both words were presented in black ink and the instructions were to remember these pairs for a later recognition test. Emotion words were adopted because the literature (e.g. Knickerbocker &

Altarriba, 2013) has reported the stronger behavioural effects with emotion words than emotion-laden words. Their performance on the subsequent recognition test was analysed to determine which type of word (positive, negative, neutral) would be appropriate to use as an abstract colour-related words. Based on the literature of emotional words and memory (e.g. Pierce & Kensinger, 2011), studying the pairs of colour names and negative emotion words was expected to reveal poorer memory accuracy than the pairs of colour names and neutral emotion words (i.e. interference) because negative stimuli tend to capture attentional resources (Pratto & John, 1991), which could result in the unequal allocation of attention to both words. Therefore, whilst the aim of the present experiment was to evaluate the appropriate type of emotion words in the AMST to produce effects on memory accuracy, it was hypothesised that negative emotion words would be most likely to show such effects.

Method

Participants

Thirty students of University of Kent were recruited as the participants. Since the task was to remember word-word pairs and Experiment 1-4 were conducted approximately half a year before this experiment, the prior participation experience for Experiment 1–4 was not considered. One male participant (age of 18) and 29 female participants (*M* = 19.97 years old; *SD* = 2.76) were recruited for the present experiment. These 30 participants consisted of 20 native English speakers, six non-native English speakers and four “English plus other language(s)” speakers.

Design

The research design was identical to Experiment 3 and 4 except the following points. In the present research the participants were asked to remember pairs of colour names (top row) and emotion words (bottom row) whilst all words were printed in black ink. There were three types of emotion words (i.e. positive, negative and neutral). The present research was a within-participants design with emotional word condition as an independent variable. The order of emotional word conditions was counterbalanced based on Williams Latin Square Design (Williams, 1949).

Materials

The practice section included new practice words. Since the task was to remember word-word pairs in the present experiment, the participants practiced with four word-word pairs in the practice section. Based on *Affective norms of English words (ANEW)* database (Bradley & Lang, 1999), the words that were emotionally neutral and had similar letter length as the four practice words in Experiment 3 and 4 were selected. The new practice correct (target) and incorrect (distractor) pairs are summarised in Appendix E. The original four practice words were presented in the top row and the new four practice ones were shown in the bottom row at the centre of the screen.

In the study and test sections the participants saw the same eight colour names as Experiment 1 and 3. In addition to it, 24 emotion words (8 words x 3 emotion word conditions) were selected based on Sutton and Altarriba's (2016) study, *ANEW* database (Bradley & Lang, 1999), Warriner, Kuperman and Brysbaert's (2013) affective ratings for English words and Brysbaert, Warriner and Kuperman's (2014) word concreteness values (see Appendix F). All words were printed in black ink throughout this experiment. Sutton and Altarriba (2016) define emotion words (e.g., sad, nice) as direct representations of emotions whilst emotion-laden words (e.g., ugly, romantic) are indirect representations of emotions although these words still induce certain emotions. The present research employed only "emotion words" for its stronger effects on task performance (e.g. Knickerbocker & Altarriba, 2013).

Positive and negative emotions words were obtained from Sutton & Altarriba's (2016) list of emotion words. Since their list showed the valence and arousal values based on the affective norm database (i.e. ANEW; Bradley & Lang, 1999), it was employed. The goal of this stimulus selection was choosing the abstract colour-related words that are identical to the colour names used in Experiment 1 and 3. The emotion words were selected through three selection processes: letter length, valence value and arousal value filtering processes. First, the words that have the identical letter length to Chapter 2 studies were selected. Then these words were filtered based on valence values in the affective norms database [i.e. ANEW and Warriner et al.'s (2013) affective ratings]. Both databases used the 9-point scale whilst lower numbers indicate low on both valence and arousal dimensions. To select positive and negative emotion words, the words that have the scores as close as possible to the extreme values were selected for each type. Finally, amongst these words, the words that show the close-to-middle arousal score were selected. This was performed due to Sutton and Altarriba's (2016) definitions. As Sutton and Altarriba's (2016) definitions implied, the category of emotion words tends to be determined based on valence values. This could indicate that the valence aspect might contribute to the substantial part of emotional impression from emotional words; the valence values might determine how people feel from reading emotional words. Consequently, the present thesis focuses on the effects from valence and excludes the intervention with arousal. Thus, after valence-based filtering, the words were also selected to show the mean arousal scores as close as possible to the moderate level of arousal (i.e. the score of 5).

Neutral emotions words were selected from Warriner et al.'s (2013) list. Since Sutton and Altarriba's (2016) study and the based database showed the limited number of neutral emotion words, this type of words was obtained from a different word list. The goal of selecting the neutral emotion words was choosing the words that were identical to the positive and negative emotion words but different only on the valence dimension. The neutral emotion words were selected through four selection processes: valence value, letter length, arousal value and word frequency selection processes. First, the words that show the approximately middle scores on the valence dimension were selected since the words should be emotionally neutral. Amongst these, the words that have the identical letter length to Chapter 2 studies were chosen. Then the words that show the approximately middle score on the arousal dimension were selected. Since these neutral emotion words needed to be abstract, the word concreteness values were examined based on the word concreteness database (Brysbaert et al., 2014). This database employed the 5-point scale and lower numbers indicate abstract words. Since the word concreteness values of the selected positive and negative words were low (Positive = 2.24 and Negative = 2.39), the low-value words were selected. Finally, word frequency was examined with SUBTLEX-UK Corpus (van Heuven, Mandera, Keuleers, & Brysbaert, 2014). This corpus employed the 7-point scale and lower numbers indicate low frequency words. As the databased defined high frequency words as having the score of 4 or higher, the present experiment attempted to select the words with the word frequency value around 4 or higher. The details of all emotion words were summarised in Table 3.1.1.

Table 3.1.1

Valence Values, Arousal Values, Word Frequency, Concreteness, Amounts and Frequency of Orthographic Neighbours and Word Length in Each Emotion Word Condition (Experiment 5-A)

	ANEW		Affective Ratings		SUBTLEX-UK	Concreteness	Orthographic Neighbours			Excel
Condition	VAL	ARO	VAL	ARO	FRE (Zipf-values)	CON	NUM	ON FRE	FRE	LEN
	M (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)
Positive	7.90 (0.49)	5.75 (1.32)	7.71 (0.49)	5.09 (1.12)	4.70 (0.94)	2.24 (0.28)	4.75 (5.23)	21.71 (45.76)	127.72 (165.31)	5.50 (1.20)
Negative	2.19 (0.41)	5.87 (1.31)	2.46 (0.34)	5.25 (0.78)	4.22 (0.39)	2.39 (0.38)	3.88 (5.74)	22.64 (36.23)	45.88 (48.00)	5.50 (1.20)
Neutral	—	—	5.00 (0.00)	4.91 (0.47)	3.65 (0.62)	2.92 (0.88)	3.00 (4.96)	13.73 (22.29)	11.96 (15.45)	5.50 (1.20)

Note. VAL = valence values; ARO = arousal values; FRE = word frequency values; CON = concreteness values; NUM = the number of orthographic neighbours; ON FRE = the averaged frequency (per million presentations of text) of orthographic neighbours; LEN = word length. These values were obtained from ANEW database (Bradley & Lang, 1999), the affective ratings (Warriner et al., 2013), SUBTLEX-UK corpus (van Heuven et al., 2014), the concreteness ratings (Brysbaert et al., 2014), MCWord database (Medler & Binder, 2005) and a Microsoft Excel (Microsoft Office Professional Plus 2016, Version 16.0.4927.1000) spreadsheet.

The positive emotion words are listed here in the alphabetical order: *admired*, *brave*, *happy*, *joyful*, *kind*, *love*, *passion* and *secure*. The negative emotion words are the followings in the same order: *cruel*, *fear*, *hostile*, *lonely*, *misery*, *rage*, *unhappy* and *upset* and the neutral emotion words are listed here in the same form: *auto*, *bleep*, *depend*, *midair*, *stretch*, *surge*, *tighten* and *toss*. These emotion words were randomly paired with colour names to create the correct pairs and incorrect pairs via the website (Randomness and Integrity Services Ltd, n.d.).

To focus on the effects between the three types of words that show different valence values, an irrelevant factor needed to be identical across these word types. Therefore, the number of orthographic neighbours of the selected emotion words was also considered. The number of orthographic neighbours represents how many same-length words are different by one letter (Madan, Scott, & Kensinger, 2019; Medler & Binder, 2005). Medler and Binder (2005) explained this concept by the example that the words *bat*, *fat*, *mat*, *cab*, etc. could be orthographic neighbours for the word *cat*. Thus, the number of orthographic neighbours represent how similar each word is. As the words have more orthographically similar words, stronger inhibition of these irrelevant information is required (Vergara-Martínez & Swaab, 2012). Thus, if the negative emotion word condition shows a higher number of orthographic neighbours and poorer memory accuracy than the positive and neutral emotion word conditions, this might be because such a high number of similar words requires additional cognitive resources; thus, it would be difficult to make a conclusion about how word concreteness affects memory accuracy with the AMST. To prevent this, this factor and the mean word frequency of the orthographic neighbours per million words for the selected words were reported following the literature (Madan et al., 2019). In this experiment the reported numbers of orthographic neighbours are the number of orthographic neighbours within MCWord database (Medler & Binder, 2005).

The goal of these emotion word selection was that each emotion word condition differed only on the valence dimension; all conditions needed to be identical in terms of arousal level, word frequency, concreteness and orthographic neighbours number. The one-way ANOVAs with the emotion word condition as a between-participant factor were conducted to analyse the difference of these properties between the three emotion word conditions. Since *ANEW* database (Bradley & Lang, 1999) does not show the valence and arousal values for the selected neutral emotion words, only the positive and negative emotion words were compared with these values in this database first. This analysis revealed that these two types of emotion words were significantly different on the valence dimension, $F(1, 14) = 633.69$, $p < .001$, partial-$\eta^2 = .98$, but not on the arousal dimension, $F(1, 14) = 0.03$, $p = .861$, partial-$\eta^2 = .002$, based on *ANEW* database values. This confirmed that the positive emotion words had the higher mean valence value ($M = 7.90$, $SD = 0.49$) than the negative emotion words ($M = 2.19$, $SD = 0.41$). The analysis also indicates that the selected positive and negative emotion words possessed the same level of arousal (positive: $M = 5.75$, $SD = 1.32$; negative: $M = 5.87$, $SD = 1.31$). The second ANOVA analyses compared all properties [i.e. the valence values based on Warriner et al.'s (2013) affective rating, word concreteness (Brysbaert et al., 2014), word frequency (van Heuven et al., 2014) and the number of orthographic neighbours and the mean word frequency of these neighbours per million words (Medler & Binder, 2005)] between each emotion word condition. Note that the homogeneity of variances were violated with Warriner et al.'s (2013) valence values and word concreteness values. The analyses also showed the significant difference between the selected emotion words on the valence dimension, $F(2, 21) = 469.50$, $p < .001$, partial-$\eta^2 = .98$.

Planned contrast confirmed that the positive emotion word possessed the higher mean valence value (*M* = 7.71, *SD* = 0.49) than the neutral emotion words (*M* = 5.00, *SD* = 0.00), *t*(7.00) = 15.73, *p* < .001, *d* = 7.86, and the mean valence value of the negative emotion word (*M* = 2.46, *SD* = 0.34) was lower than the neutral emotion words, *t*(7.00) = -21.21, *p* < .001, *d* = 10.61. For other properties except word frequency, there was no difference between the three types of the selected emotion words [Arousal: *F*(2, 21) = 0.33, *p* = .723, partial-η^2 = .03; Word concreteness: *F*(2, 21) = 3.08, *p* = .067, partial-η^2 = .23; The number of orthographic neighbours: *F*(2, 21) = 0.22, *p* = .807, partial-η^2 = .02; Frequency of orthographic neighbours per million presentations of text: *F*(2, 21) = 0.15, *p* = .864, partial-η^2 = .01]. For word frequency, the analysis revealed the significant difference, *F*(2, 21) = 4.75, *p* = .020, partial-η^2 = .31. Planned contrast indicated that the mean word frequency in the positive emotion word condition (*M* = 4.70, *SD* = 0.24) was higher than the neutral emotion word condition (*M* = 3.65, *SD* = 0.24), *t*(21) = 3.08, *p* = .006, *d* = 1.33, and there was no difference between the selected negative emotion words (*M* = 4.22, *SD* = 0.24) and the selected neutral emotion words, *t*(21) = 1.66, *p* = .111, *d* = 1.10, and between the positive and negative emotion words, *t*(21) = 1.42, *p* = .171, *d* = 0.68. Whereas SUBTLEX-UK corpus (van Heuven et al., 2014) word frequencies revealed a significant difference between the positive emotion words and the neutral emotion words, this difference disappeared with the word frequencies from MCWord database (Medler & Binder, 2005), which revealed word frequency per million presentations of text, *F*(2, 21) = 2.85, *p* = .081, partial-η^2 = .21 (Positive: *M* = 127.72, *SD* = 165.31; Negative: *M* = 45.88, *SD* = 48.00; Neutral: *M* = 11.96, *SD* = 15.45). Due to this instability of word frequency differences between the selected emotion words across the databases, it would be impractical to explain memory performance differences based on word frequency; the differences of this value were ignorable in the present experiment. Since the selected 24 emotion words were different only on the valence dimension and were identical on other features except this word frequency measure, these words were employed in this experiment.

For the format of presenting stimuli, one might question why presenting word-number pairs as in the previous chapter or using pairs of emotion words and colour patches may be inappropriate in the present experiment. This was because the former case would fail to activate the semantic relationships between items and the latter would have the issue of confirming semantic relationships between emotion words and colour patches. For example, presenting un*happy-7* with black ink would be difficult to activate semantic relationships in the AMST. To trigger such relationships, the information presented on the AMST needs to be semantically associated. Since this experiment did not include colour-word associations yet in order to reduce the complexity of research, such associations need to exist between the items on the AMST. Thus, it would be necessary to examine how strongly the selected colour names and emotion words are semantically related. Such semantic relationships between colour names and emotion words were confirmed by Latent Semantic Analysis (Landauer & Dumais, 1997; for an overview see Landauer, Foltz, & Laham, 1998). This is the theory that helps researchers to assess how words share contextually similar meanings to each other. The indices of semantic similarity between each colour name and emotion word was computed by the website (Laham & Steinhart, 1998) in the format of matrix comparison (e.g. the word *blue* was compared with each of eight positive emotion words, of eight negative emotion words and of eight neutral emotion words) at the level of general reading for the first year college. This reading level was selected since this was the only option that is set for college level. The larger index indicates a more contextually-similar meaning shared between words (the maximum value is 1.00). For example, the website (Laham & Steinhart, 1998) demonstrated that the similarity index between the word *cat* and *mouse* was 0.42. The semantic similarity between the selected eight colour words and each type of emotion words were: the positive emotion word condition (*M* = .31, *SD* = .08), the negative emotion word condition (*M* = .33, *SD* = .05) and the neutral emotion word condition (*M* = .24, *SD* = .11). The one-way ANOVA with emotion word type as a between-participant factor revealed that there was a significant difference between these three emotion word conditions, $F(2, 45) = 4.39$, $p = .018$, partial- $\eta^2 = .16$. Independent sample *t*-test (two-tailed) showed that the semantic similarity indices did not differ between the positive and negative emotion word conditions [$t(30) = -.078$, $p = .542$, $d = 0.28$]. However, independent sample *t*-tests (one-tailed) revealed that the colour names showed the stronger semantic similarity with the positive emotion words than with the neutral emotion word [$t(26.738) = 1.92$, $p = .0325$, $d = 0.68$] and the same pattern was observed with the negative emotion words than with the neutral emotion words [$t(21.155) = 2.72$, $p = .0065$, $d = 0.96$]. Thus, the analyses confirmed that using these selected emotion words were likely to activate semantic relationship information with colour names, which seemed to function in the identical process as Experiment 1 and 3 of the present thesis. Since Latent Semantic analysis can be performed with words, this analysis is unable to assess such semantic relationships between emotion words and colour patches; this could imply the difficulty of presenting emotion words with colour patches in the present experiment. Therefore, in this experiment the pairs of colour names and emotion words were presented with black ink. As in Experiment 1-4 in the present thesis, the fixed word-word pairs were used across all participants.

Procedure

The entire procedure was identical to Experiment 3 and 4. The main modification from these two previous experiments was that the numbers were replaced with the emotion words and both colour names and emotion words were written in black ink throughout this experiment. Additionally, for the future analysis, the present experiment included the emotional state questionnaires in the beginning and at the end and the word type questionnaire was also inserted at the end. After the participants completed the informed consent and saw the title screen, they were asked to answer two questions about their current mood by typing the corresponding numbers. Then the message, “How are you feeling now?”, and the picture of 9-point Likert scale were shown on the screen. First, the pleasant scale was shown with *1* represented “unpleasant” and *9* referred to “pleasant.” After they typed the number, this scale disappeared and the arousal scale appeared on screen that *1* represented “calm” and *9* indicated “excited.” The message was presented at the top part of both screens and the scales were shown below that. Both message and pictures of the scales were permanently presented until the participant typed the corresponding numbers, which were on the top row of the QWERTY keyboard. Upon completion of these two questions, the participants saw the task instruction and followed the same research procedure as Experiment 3 and 4 in the present thesis.

After the participants finished the third emotion word condition block, they were asked to answer the same two-trial emotion questionnaire. When they completed this, the participants experienced the one-trial word type questionnaire to examine whether they had noticed that this experiment included three types of emotion words. During this word type questionnaire, the participants were asked which type of emotion words they had studied. They answered this questionnaire by typing the corresponding number. The numbers are listed here: 1 (*positive + negative + neutral*), 2 (*positive + negative*), 3 (*positive + neutral*), 4 (*negative + neutral*), 5 (*only positive*), 6 (*only negative*), 7 (*only neutral*) and 8 (*I don’t remember*). These response key options kept appearing until the participants made a response. Note that there was no time limit when answering all questionnaires. Therefore, there were 4 [(valence + arousal) x 2 (pre-test + post-test)] emotion state trials + 48 [(8 correct word-word pairs + 8 incorrect pairs) x 3 emotion word conditions] recognition memory test trials + 32 [(8 same pairs + 8 different pairs) x 2] tone-comparison trials + one-trial word type questionnaire in the present experiment. The present experiment took approximately 15 minutes to complete.

Results

As in Experiment 3 and 4, mean hit (Hit) rates, false alarm (FA) rates, d' values, c values and reaction times in each emotion word condition are summarised in Table 3.1.2 and Figure 3.1.

Table 3.1.2

Hit Rates, False Alarm Rates, d' Values, c Values and Reaction Times (in ms) in Each Emotion Word Condition (Experiment 5-A)

Condition	Hit Rate	FA Rate	d'	c	Reaction Time
	M (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)
Positive	.81 (.14)	.23 (.17)	1.89 (1.01)	-0.06 (0.27)	1,743 (413)
Negative	.74 (.17)	.31 (.19)	1.35 (1.04)	-0.10 (0.31)	1,958 (528)
Neutral	.81 (.11)	.25 (.16)	1.79 (0.96)	-0.09 (0.24)	1,647 (423)

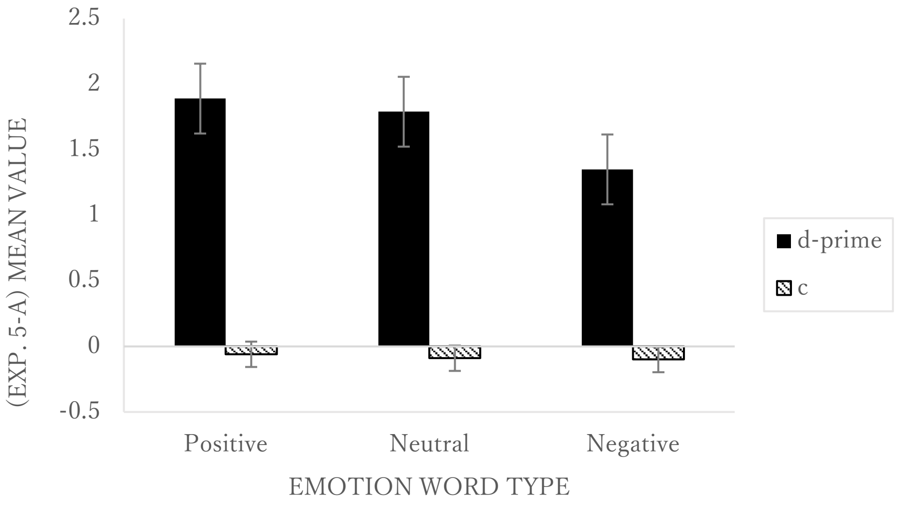


Figure 3.1. The graph shows the mean d' values and the mean c values in each emotion word condition (Experiment 5-A). Error bars represent the 95% confidence intervals (Masson & Loftus, 2003).

None of the participants was timed out and all participants' response times in each trial were longer than 200 ms, which indicated their appropriate attention to the task. Therefore, all responses were included in the following analyses.

Analysis of Accuracy

The one-way ANOVA with Emotion (Positive, Negative and Neutral) as a within-participant factor was conducted (see Table 3.1.2 and Figure 3.1). The analysis of d' values revealed a significant main effect of Emotion, $F(2, 58) = 4.59$, $p = .014$, partial-$\eta^2 = .14$. Paired sample *t*-tests (two-tailed) indicated that the negative emotion word condition (M = 1.35, SD = 1.04) showed poorer memory accuracy than the neutral emotion word condition (M = 1.79, SD = 0.96), $t(29) = -2.31$, $p = .028$, $d = 0.42$, and than the positive emotion condition (M = 1.89, SD = 1.01), $t(29) = -2.98$, $p = .006$, $d = 0.54$. There was no significant difference between the positive emotion condition and the neutral emotion word condition, $t(29) = 0.52$, $p = .609$, $d = 0.09$.

In terms of response bias (i.e. c values), the main effect of Emotion was not significant, $F(2, 58) = 0.22$, $p = .805$, partial-$\eta^2 = .01$, and all three emotional word conditions revealed identical negative c values (Positive: M = -0.06, SD = 0.27; Negative, M = -0.10, SD = 0.31; Neutral, M = -0.09, SD = 0.24), which replicated the findings of Experiment 3 and 4 (i.e. the weak tendency to answer "old" to all test items).

Analysis of the Emotional State Questionnaires

The participants' mean scores on the emotional state questionnaires (i.e. asking them to give a rating on the 9-point pleasure scaling and arousal scaling) were compared between pre-test and post-test. Reaction times on these questionnaires were also longer than 200 ms, which indicated that the participants responded to these questions with the appropriate attention. Paired sample *t*-tests (two-tailed) revealed that the participants' valence rating significantly decreased from pre-test (M = 6.63, SD = 1.38) to post-test (M = 6.07, SD =

1.57), $t(29) = 2.73$, $p = .011$, $d = .50$, whilst the arousal scores did not change (pre-test: M = 5.57, SD = 1.63; post-test: M = 5.27, SD = 1.55), $t(29) = 1.07$, $p = .293$, $d = .20$. Although there was a significant reduction on the valence scores, the 9-point scales consisted of only whole numbers ranging from 1 to 9; therefore, both pre-test and post-test mean scores were still the score of 6 on the 9-point scale. This could indicate that the participants were feeling the moderate level of pleasure throughout the experiment whilst the slight reduction of pleasure scores might be explained by the aversive feelings due to the mental fatigue (for an overview see Boksem & Tops, 2008). Therefore, this stability of emotional state confirmed that the reported memory performance was not affected by changes in the participants' mood during the task.

Analysis of Awareness

To examine whether the participants' awareness of emotion word types modified recognition memory performance, they were categorised into two groups: those who correctly recognised the emotion word types (i.e. those who answered 1 on the one-trial word type questionnaire) and those who did not (i.e. those who selected other numbers on the questionnaire). With d' values, the 3 (Emotion) x 2 (Word Awareness) factorial ANOVA was conducted with Emotion as a within-participant factor and Word awareness as a between-participant factor whilst 12 participants correctly recognised the emotion word types. Since the analysis revealed that the interaction effect of Emotion and Word Awareness was not significant [$F(2, 56) = 0.76$, $p = .474$, partial-$\eta^2 = .03$], this indicated that whether the participants correctly realised that they had studied three types of emotion words or not had little impact on memory performance.

Discussion

Overall, the results indicated that the participants showed the poorest recognition memory accuracy with the pairs of colour names and negative emotion words. Thus, evidence from the analyses supported the hypothesis, which expected effects on memory accuracy from emotion words, especially from negative emotion words.

The observation of difference in memory performance with the abstract colour-related words (i.e. the emotion words) could bolster the mental imagery account (i.e. since colour-related words could induce concrete mental imagery, it aided the participants to associate the colour-related words and incongruent ink colours, which led to no ink colour effects on memory accuracy). However, the poorer memory accuracy with the negative emotion words might raise a question; if word concreteness modifies the robustness of ink colour effects on the AMST, why only negative emotion words revealed memory degradation. The answer for this question could be provided by Brainerd, Stein, Silveira, Rohenkohl and Reyna's (2008) study of negative emotional words and false memories.

The researchers created the lists that consisted of three types of emotional words: positive, negative and neutral. They employed a recognition test with three types of distractors (i.e. critical, semantically related and unrelated distractors) to test how affective valence of written words contribute to the occurrence of false memories. Critical distractors referred to the words that were semantically related to the critical non-studied words whilst semantically related distractors were the unpresented but same affective-valenced words and unrelated distractors represented the collection of the non-studied words that were retrieved from all three types of emotions. Brainerd et al. assumed that the response accuracy to the critical distractors would reflect false memories. The clear pattern of the results emerged; false memory occurred more with the negative critical distractors than the neutral ones and with the neutral critical distractors than the positive ones. Additionally, further analyses indicated that increased meaning similarity between the presented and unpresented items contributed to the higher occurrence of false memories with negative-valenced stimuli. Their conclusion that semantic similarity between the target and distractors affect memory accuracy can provide an explanation for the present results (i.e. the poorest memory accuracy in the negative emotion word condition whilst no facilitation occurred with the positive emotion words). In the current experiment Latent Semantic Analysis (Landauer & Dumais, 1997; for an overview see Landauer, Foltz, & Laham, 1998) confirmed that the semantic similarity between the positive emotion words and the colour names (e.g. the word *brave* and the colour name *blue*) and between the negative emotion words and the colour names (e.g. *upset* and *green*) were significantly stronger than between the neutral emotion words and the colour names (e.g. *surge* and *brown*). Thus, if any task performance was observed with the emotion words, semantic similarity could be a potential factor. The main finding of the present experiment was that the negative emotion word condition revealed the poorest memory accuracy and there was no difference between the positive emotion word condition and the neutral emotion word condition. Since semantic similarity between the items becomes influential with negative stimuli, not with positive stimuli, based on the study conducted by Brainerd et al. (2008), such a factor appeared to lead the present experiment to reveal the reported results; only negative stimuli had an impact on memory accuracy.

Alternatively, our results could also be explained by the perspective of unequal attentional distribution due to negative stimuli. The interesting feature of the reported results is that these appear to contradict the prevailing notion regarding negative stimuli. Based on the results in the present experiment, it would be reasonable to question why this study revealed poorer memory accuracy with negative emotional stimuli whilst it has been reported that negative events tend to induce vivid memories (e.g. flashbulb memories; Reisberg & Heuer, 2004). Despite such a common pattern of finding, Reisberg and Heuer highlighted the fact that memory accuracy can be impaired due to emotional factors. For instance, they noted that increased attention to emotional stimuli could contribute to the degradation of memories for surrounding information. This unequal attentional distribution was also bolstered by Pratto and John’s (1991) study, which found that negative stimuli attracted more attention than positive stimuli. Therefore, the poorest memory accuracy with the negative emotion stimuli could be because the participants’ spent longer time on the negative emotion words and paid less attention to the paired colour names. With either account, it was found that the negative emotion words would be more likely to show effects when effects of task-irrelevant information on the AMST are measured by memory tests. Thus, the next experiment used this type of abstract colour-related words and examined whether this attentional account would be more feasible than word concreteness hypothesis to explain memory impairment with emotion words.

Experiment 5-B

Experiment 5-B examined whether increasing word abstractness induces effects of task-irrelevant information on memory performance when the words on the AMST are negative emotion words. Since Experiment 5-A found poorer memory accuracy with negative emotion words, this could provide support for word concreteness account and could indicate that this type of colour-related words might produce such effects when task-irrelevant semantic relationships between ink colours and words are manipulated. Therefore, the present experiment adopted this type of abstract colour-related words in the AMST. Whilst Experiment 5-A revealed interference with negative emotion words, this could be due to semantic similarity between the colour names and the negative emotion words because such a semantic feature tends to induce false memories with negative stimuli rather than positive ones (Brainerd et al., 2008). Alternatively, unequal attentional allocation might have produced the impaired memory with the negative emotion words since it has been found that negative stimuli tend to attract participants' attention more (Pratto & John, 1991). However, this would be reasonable only when comparing different categories of emotional stimuli (e.g. positive stimuli and negative stimuli). By adopting the same type of emotional stimuli, if memory performance is significantly different across conditions, it can be expected that this attentional account will become no longer cogent since all conditions suffer from such an attentional control issue. Therefore, adoption of a specific type of emotional word will allow researchers to focus on different intrinsic features of stimuli such as semantics of words.

Based on this assumption, the present experiment explored the issue; whether using abstract colour-related words produce effects when performance on the AMST is measured by memory tests. Experiment 5-B aimed to test whether task-irrelevant ink colours printed in numbers and its relationship with abstract colour-related words (i.e. negative emotion words) in the AMST affect recognition memory accuracy of word-number pairs. To achieve this aim, the participants were first instructed to select their personal image colours for the presented emotion words. After their memory of which colours they had chosen was tested, they were asked to study the pairs of emotion words and numbers for the later recognition test and accuracy of their responses on this memory test was analysed. The hypothesis was that when ink colours printed in numbers were incongruent with the personal image colours of negative emotion words (e.g. *unhappy-2* with the word *unhappy* was presented in black ink but the number *2* was written in yellow ink when the participant selected the colour “blue” as the personal image colour for *unhappy*), recognition memory test scores would be lower than when both items were presented in a neutral ink. Since this interference effect has been consistently reported from the previous chapter, the present experiment expected to observe only this type of effect; no facilitation was expected. If such interference was observed, it might be possible to bolster the hypothetical account that no ink colour effects with colour-related words in Chapter 2 occurred because the participants could link the words and numbers (which were printed in incongruent ink colours) due to the effortless processes of creating mental pictures (e.g. to remember the *strawberry-7* pair with the number *7* in white ink, the participants could imagine seven strawberries on top of whipped cream) with concrete words (Paivio, 1965, 1991, 2006; Paivio & Csapo, 1969; Paivio et al., 1966; Yarmey & Paivio, 1965). In other words, it was expected that by reducing concreteness of colour-related words, learners might experience difficulty of combining words and numbers with incongruent ink colours that might result in poorer memory accuracy. The present experiment tested this assumption.

Method

Participants

Thirty students of University of Kent were recruited as the participants. Their prior experience of participating in the previous experiments (i.e. Experiment 1-4) was not considered as Experiment 5-A since there was at least half-a-year interval between these experiments and the present experiment whilst the present experiment was restricted to those who did not participate in Experiment 5-A. Two male participants (*M* = 19.50 years old; *SD* = 0.71) and 28 female participants (*M* = 19.75 years old; *SD* = 3.06) were recruited for the present experiment. These participants consisted of 25 native English speakers, three non-native speakers and two “English plus other language(s)” speakers.

Design

The research design was identical to the previous experiment except where noted. There were three main modification from Experiment 5-A: different format of to-be-studied materials, different manipulation of ink colour congruency and the additional post-test questionnaire. First, the research design returned to the standard AMST that showed the negative emotions words on top and the numbers below. To manipulate semantic congruency between task-irrelevant ink colours and the emotion words, this was set up individually by using the participants’ personal image colours for each emotion word as congruent ink colours. As the final difference, the present experiment added the ink colour congruency check questionnaire, which aimed to test whether the participants had noticed the ink colour difference, at the end of the experiment.

Materials

Differing from Experiment 1-5.A, white ink was used as a neutral ink colour. The literature (Sutton & Altarriba, 2016) demonstrated that negative emotion words tend to trigger the colour of black rather than of white. Since this would suggest stronger semantic relationships would occur with black ink than white ink, black was used as an ink colour of numbers in the AMST.

This experiment used the same negative emotion words as the previous experiment (see Appendix G). After the participants experienced the pre-test emotion questionnaire, they were asked to select their personal image colours for the individually presented negative emotion words (see Figure 3.2.1). The words were presented in the same ink colour, font and size as the to-be-studied words in the study section. The width of these coloured circles was 0.08 points and the height was 0.1 points in Psychopy (Peirce et al., 2019). The ink colours of these circles were the same ink colours that were used in Experiment 1-4 and RGB code for the colour “black” was (0, 0, 0).

Figure 3.2.1. The example screen of the personal image colour selection section.

The materials in the practice, study and test sections were the same as Experiment 3 and 4 except the use of the negative emotion words and the following point. To create correct and incorrect word-number pairs, the same correct-incorrect number combinations were used as in Experiment 4 (e.g. in the congruent condition the number *5* was used in the correct pair and the number *2* was presented in the incorrect pair). These correct-incorrect number combinations were randomly paired with the selected eight negative emotion words.

After the participants answered the post-test emotion questionnaire, they were instructed to answer the one-trial ink colour congruency check questionnaire at the end of the experiment. This questionnaire was inserted to examine whether semantic congruency between ink colours and the negative emotion words was successfully activated during study. To clarify the definitions of congruent, incongruent and neutral ink colours, the participants saw the examples before the questionnaire [e.g. “In the case that you selected 1. ● (blue-coloured circle) for *unhappy*, if you studied the *unhappy-3* pair (the number *3* was written in blue ink) → ‘congruent’, the *unhappy-3* pair (*3* in red ink) → ‘incongruent’ and the *unhappy-3* pair (*3* in white ink) → ‘neutral’”]. On this questionnaire, the instruction (i.e. “Which ink colour type(s) did you see in word-number pairs?”) was permanently presented on top of screen and the participants were asked to type the appropriate response key without the time limit. The response options were presented below the instruction and are listed here: 1 (*consistent + inconsistent + neutral*), 2 (*consistent + inconsistent*), 3 (*consistent + neutral*), 4 (*inconsistent + neutral*), 5 (*only consistent*), 6 (*only inconsistent*), 7 (*only neutral*) and 8 (*I don’t remember*). These response options were also kept visible on screen until the participants made a response.

Procedure

The procedure was identical to Experiment 3, 4 and 5-A except where noted. After completing the two-trial emotional state questionnaire, the participants were asked to choose a personal image colour for the presented emotional word. Note that they could select the same colours as many times as they would like to and they were informed to remember their selection for the later test. Upon the completion of this selection, the eight-trial forced-choice recognition test (i.e. a test that asks participants to select the target stimuli amongst distractors; Stanislaw & Todorov, 1999) was inserted to enhance the participants’ memory of these personal colour-word associations. The design of this forced-choice recognition test screen was identical to the previous section (i.e. the personal-image-colour selection section) whilst the participants were asked to choose the same personal image colours that they had selected in the previous section for each of the presented negative emotion words. For instance, if the participant selected 1. blue-coloured circle for the word *unhappy*, this person was required to type the "1" key when this negative emotion word appeared on screen. During the forced-choice recognition test, the feedback was provided for 2,000 ms and it presented the correct answer when they made an incorrect response. In the case of the *unhappy*-"blue" pair, when this participant made an incorrect response key, the message "Oops! That was wrong! Correct Answer = *blue*" appeared. In this message the previously-selected colour was presented in the written word format (e.g. *blue*, *red*) with white ink. Since the feedback message was longer than Experiment 1-5.A, the feedback presentation duration was increased in this experiment. This forced-choice recognition test repeated three times without the time limit. After the third forced-choice recognition test, the participants followed the same procedure as Experiment 3, 4 and 5-A. The incongruent ink colours were randomly selected from the rest of seven ink colours (e.g. if the participant selected the colour "blue" as the personal image colour for the word *unhappy*, in the incongruent condition this person saw the *unhappy*-*2* pair with the number *2* was written in red ink). When they finished the post-test two-trial emotional state questionnaire, they were asked to answer the ink colour congruency check questionnaire. Thus, this experiment consisted of 4 [(valence + arousal) x 2 (pre-test + post-test)] emotional state questionnaire trials + 24 (8 negative emotion words x 3 sessions) forced-choice recognition test trials + 48 [(8 correct + 8 incorrect) x 3 colour congruency conditions] recognition memory test trials + 32 [(8 same pairs + 8 different pairs) x 2] tone-comparison trials + 1 ink colour congruency check questionnaire trial. The experiment took approximately 15 minutes to complete.

Results

As in Experiment 3, 4 and 5-A, mean hit (Hit) rates, false alarm (FA) rates, d' values, c values and reaction times in each colour congruency condition are summarised in Table 3.2 and Figure 3.2.2.

Table 3.2

Hit Rates, False Alarm Rates, d' Values, c Values and Reaction Times (in ms) in Each Colour Congruency Condition (Experiment 5-B)

Condition	Hit Rate	FA Rate	d'	c	Reaction Time
	M (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)
Congruent	.66 (.18)	.38 (.18)	0.83 (0.99)	-0.05 (0.26)	2,102 (562)
Incongruent	.66 (.20)	.36 (.17)	0.87 (0.99)	-0.02 (0.32)	2,197 (836)
Neutral	.64 (.15)	.40 (.16)	0.69 (0.78)	-0.07 (0.28)	2,175 (509)

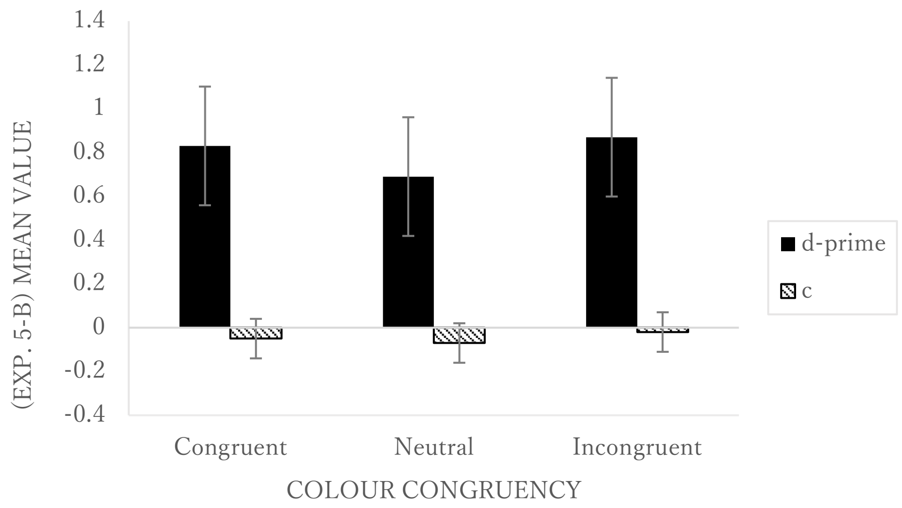


Figure 3.2.2. The graph shows the mean d' values and the mean c values in each colour congruency condition (Experiment 5-B). Error bars represent the 95% confidence intervals (Masson & Loftus, 2003).

Analysis of the Forced-Choice Tests

On the forced-choice tests, the participants' reaction times in all trials were longer than 200 ms, which signified their appropriate attention. To confirm that the participants' learning accuracy of colour-word associations was enhanced by repeating the sessions, the one-way ANOVA with Forced-choice sessions as a within-participant factor was conducted. This analysis revealed a significant main effect of Forced-choice sessions, $F(2, 58) = 5.26$, $p = .008$, partial-$\eta^2 = .15$. Paired sample *t*-tests (two-tailed) indicated the significant increment of the participants' correct responses on the forced-choice test from the first session (*M* = .85, *SD* = .17) to the second (*M* = .91, *SD* = .13), $t(29) = 3.00$, $p = .005$, $d = 0.55$. Their test scores did not significantly change from the second to the third (*M* = .93, *SD* = .10), $t(29) = 0.50$, $p = .620$, $d = 0.09$. However, when comparing the first with the third, there was the significant increment of the test scores, $t(29) = 2.57$, $p = .016$, $d = 0.47$. This confirmed that the participants' learning of colour-word associations was enhanced by the three sessions of the forced-choice recognition tests. No occurrence of speed-accuracy tradeoff on the forced-choice tests was also confirmed by the analyses.

Analysis of Accuracy

On the recognition tests for word-number pairs, none of the participants was timed out and 29 participants' response times in each recognition memory trial were longer than 200 ms, which indicated their appropriate attention to the task. Whilst one participant responded in less than 200 ms, this participant made this quick response only in one trial out of 48 trials that also could signify this person's appropriate attention to the task. Therefore, all participant's responses were included in the analyses. The same one-way ANOVA with Colour congruency as a within-participant factor was conducted (see Table 3.2 and Figure 3.2.2) as Experiment 3 and 4. The analysis with d' values revealed that there was no significant main effect of Colour congruency, $F(2, 58) = 0.47$, $p = .625$, partial-$\eta^2 = .02$. Even when comparing the congruent and incongruent conditions only, paired sample *t*-test (two-tailed) revealed no significant difference, $t(29) = -0.18$, $p = .859$, $d = 0.03$.

For response bias, the same ANOVA was conducted and also showed no significant main effect of Colour congruency, $F(2, 58) = 0.31$, $p = .734$, partial-$\eta^2 = .01$.

Whilst the analysis showed no ink colour effects on recognition memory accuracy in the present experiment, one might question whether this result occurred due to weak associations between the ink colours and the negative emotion words; the participants might not have remembered what the congruent ink colour for the word *unhappy* was during study. Since the previous ANOVA analysis included all participants who scored under 100% on the forced-choice recognition test, the re-calculation of ink colour effects only with those who revealed the perfect score on this test might modify the previously-reported outcome. To test this assumption for memory accuracy, the same one-way ANOVA was performed with 17 participants who scored 100% on the third session (i.e. the last session) of the forced-choice tests, which could indicate the participants' genuine learning of colour-word associations. However, this analysis also revealed no significant main effect of Colour congruency, $F(2, 32) = 0.46$, $p = .634$, partial-$\eta^2 = .03$. Alternatively, it would be possible to suspect whether there is a relationship between performance on the forced-choice recognition tests and recognition memory accuracy of word-number pairs (e.g. as the participants scored higher on the third session of the forced-choice tests, they might have revealed more robust effects of colour congruency or less effects due to more superior memory abilities). Therefore, the one-way Analysis of Covariance (ANCOVA) was performed with Colour congruency as a within-participant factor and The 3rd session performance as a covariance. Again, the analysis showed no significant main effect of Colour congruency, $F(2, 56) = 0.02$, $p = .977$, partial-$\eta^2 = .001$, as well as the interaction effect of Colour congruency x The 3rd session performance, $F(2, 56) = 0.05$, $p = .956$, partial-$\eta^2 = .002$. From these two analyses, it could be concluded that whether the participants remembered their personal image colours for the negative emotion words accurately or not has no impact on the robustness of the subsequent colour-word associations.

Analysis of the Emotional State Questionnaires

The participants' responses on the emotional state questionnaires (i.e. asking them to rate on the 9-point pleasure scale and the 9-point arousal scale in the beginning and end of the experiment) was analysed as the previous experiment. All reaction times in each trial were longer than 200 ms, which signified the participants' appropriate attention to the task. Paired sample *t*-tests revealed that the pleasure scores significantly decreased from the pre-test (M = 5.70, SD = 1.62) to the post-test (M = 4.93, SD = 1.57), $t(29) = 2.72$, $p = .011$, $d = 0.50$, whilst there was no differences with the arousal scores (Pre-test: M = 4.93, SD = 1.48; Post-test: M = 4.93, SD = 1.41), $t(29) = 0.00$, $p = 1.000$, $d = 0.00$. Since it was the 9-point scale, the scores from 4 to 6 would be considered as moderate whilst 1 to 3 and 7 to 9 would be extreme on both scales. Although there was a significant reduction on the pleasure scale, the mean scores were approximately 6 at the pre-test and 5 at the post-test, which would be considered as the moderate level of pleasantness. Additionally, this reduction of the pleasure value might be explained in the same line with Experiment 5-A (i.e. lower pleasure values might have occurred due to the mental fatigue; for an overview see Boksem & Tops, 2008). Thus, it was not discussed further. There were no significant reaction time differences between the pre-test and post-test on both pleasure and arousal scales (all $p > .05$).

Analysis of Awareness

To examine whether the participants' notification of colour-word congruency during study modifies the robustness of the ink colour effects, the 3 (Colour congruency) x 2 (Ink Awareness) factorial ANOVA was performed with Colour congruency as a within-participant factor and Ink colour congruency accuracy as a between-participant factor and with d' values whilst 13 participants correctly recognised all three types of colour congruency. This analysis demonstrated no interaction effect of Colour congruency x Ink awareness [$F(2, 56) = 1.12$, $p = .334$, partial-$\eta^2 = .04$], which indicated that the robustness of ink colour effects were stable regardless of the participants' notification of colour-word congruency during learning.

Discussion

Having combined these analyses, the present experiment replicated no ink colour effects with colour-related words on memory accuracy as Experiment 2 and 4 showed.

General Discussion of Experiment 5-A and 5-B

Experiment 5-A and 5-B explored whether the reduction of word concreteness by using abstract colour-related words on the AMST can overturn the previously-observed no ink colour effects with concrete colour-related words. This question was tested with using emotional words, which were expected to function as abstract colour-related words. Experiment 5-A was a preliminary experiment, which aimed to specify what type of emotional words is more likely to produce effects on memory with the AMST. In Experiment 5-A, based on the distinctive feature of negative emotional words (e.g. Pierce & Kensinger, 2011; Pratto & John, 1991), it was hypothesised that negative emotional words would be more likely to produce effects on memory performance compared to positive and neutral emotional words. To test this hypothesis, Experiment 5-A employed three types of emotion words: positive, negative and neutral. Emotion words refer to the direct label of emotion (e.g. *happy, sad*; Sutton & Altarriba, 2016) and this type of words were adopted due to its stronger emotional effects (e.g. Knickerbocker & Altarriba, 2013) on task performance. In Experiment 5-A the participants were asked to remember the pairs of colour names and emotion words for the later recognition tests. The important trait of this experiment was that all items were printed in black ink; no colour congruency was involved yet. As expected, the analyses demonstrated the poorest memory accuracy for the negative emotion words. Therefore, it was concluded that negative emotion words would be the most appropriate as an alternative of using colour-related words.

Building upon this preliminary experiment, Experiment 5-B aimed to test the main research question with using negative emotion words, which were expected to function as abstract colour-related words. The hypothesis in Experiment 5-B was similar to the previous experiments; when the ink colours printed in the numbers were incongruent with the image colours of negative emotion words (e.g. *unhappy-2* with the number *2* was written in yellow ink), this colour-word incongruency impairs the participants' recognition memory accuracy of word-number pairs. The research design was identical to Experiment 3 and 4; the participants were asked to study pairs of negative emotion words and numbers for the subsequent recognition memory tests. In the congruent condition the numbers were printed in the previously-selected personal image colours whilst the incongruent ink colours were randomly selected from the rest of seven ink colours. As in Experiment 3 and 4, the participants' correct responses on the recognition memory tests were used in the analyses. The results replicated the findings of Experiment 2 and 4; the negative emotion words failed to produce effects on memory accuracy, which did not support the hypothesis.

Although these two experiments revealed a different effect on memory, both findings could be explained by the same factor; attentional distribution. For Experiment 5-A results, this reported pattern of recognition memory accuracy matched Pierce and Kensinger's (2011) study, which revealed the poorest recognition memory accuracy with the pairs of negative-valence words after a short delay because negative emotion tends to produce narrower attention, and was in the same line with the findings of Pratto and John's (1991) research, which reported interference with reaction time in the traditional Stroop task (i.e. a colour naming task) with negative-valence stimuli due to increased attention to this type of stimuli. Having combined these studies, it could be possible to speculate that since the negative emotion words captured the large amount of the participants' attention in Experiment 5-A, this might have led them to pay less attention to the paired colour names. Thus, this difficulty of associating the colour names and negative emotion words might have contributed to the poorest memory accuracy.

Whilst the attentional control account seems to be a potential explanation for "why" interference on the accuracy occurred in Experiment 5-A, the specific processes of "how" these produce such a negative effect on memory tests could be explained from a different perspective; the cognitive load theory perspective. That is, whereas the literature (e.g. Pierce & Kensinger, 2011; Pratto & John, 1991) did not specify the underlying mechanisms of how automatic attention captured by negative-valence stimuli contributed to the interference effect, the present thesis suggests that this process could be explained by imposed extraneous cognitive load with the negative emotion words. Specifically, as Hazan-Liran and Miller (2017) referred, attentional disengagement (Posner, 1980) might have produced the reported interference on memory accuracy with the negative emotion words at the expense of cognitive resources. During study and test in the negative emotion word condition, the participants were presented the pairs of colour names and negative emotion words. Whilst the negative emotion words attracted the participants' attention, they needed to process both words (i.e. the colour names and emotion words) to learn associations between these words during study and to judge whether the pairing of these words was correct or incorrect during test. To do these, the participants were required to intentionally shift their attention from the negative emotion word to the colour name and this attentional disengagement required the suppression of task-irrelevant emotional information at the expense of scarce cognitive resources (Hazan-Liran & Miller, 2017). The present thesis proposes that this required cognitive resources can be considered as extraneous cognitive load because this use of mental energy is irrelevant to the task (i.e. learning the word-word pairs). This attentional disengagement was required only in the negative emotion word condition; consequently, only this group suffered from the higher level of extraneous cognitive load and it resulted in the poorest memory accuracy.

These two accounts of attentional capture and cognitive load hypothesis also seem to offer an explanation for no ink colour effects in Experiment 5-B. With the attentional account, the negative emotion words failed to produce ink colour effects on memory accuracy since, in all colour congruency conditions, the participants' attention might have drawn more to the negative emotion words rather than the coloured numbers, which may have waned colour-word congruency effects during study. The cognitive load hypothesis could describe such results by the tentative explanation that extraneous cognitive load was equally imposed in all colour congruency conditions by attentional disengagement processes for the negative emotion words and the effects of colour-word congruency with this type of colour words were too weak, thus, ignorable during study.

In summary, evidence from Experiment 5-A and 5-B suggests the potential conclusion; word concreteness would be inappropriate as a factor that needs to be included in the existing model of cognitive load theory. Since reducing word concreteness failed to induce effects from task-irrelevant ink colours when such effects on the AMST are measured by memory tests, it is implausible to explain why only colour names showed effects in the previous chapter from the perspective of word concreteness. Therefore, whilst this factor seems to have an impact on task performance in the memory literature (e.g. Paivio, 1965), this was not the case in the present research.

Experiment 6

The present experiment examined whether the experience of memory transformation affects the robustness of the ink colour effects on the AMST when such effects are measured by memory tests. Chapter 1 demonstrated that sleep can help paired-associate learning (Fenn & Hambrick, 2012, 2013; Potkin & Bunney, 2012) since sleep can transform short-lived memories of learned pairs into a more durable memory form (for a summary see Campos et al., 2010). Since the literature (e.g. Diekelmann & Born, 2010) has reported the importance of pre-existing schemas when transforming short-lived memories into long-lived memories, the present experiment applied this assumption to test the use of colour-related words on the AMST in memory research. The process of transferring newly-acquired information into long-term memories is named “consolidation” (Diekelmann & Born, 2010; Klinzing et al., 2019). Diekelmann and Born (2010) and Klinzing, Niethard and Born (2019) described that sleep has been reported to induce this process and to enhance newly-acquired knowledge to become more stable and durable. The interesting feature of this sleep-induced consolidation processes is its relationship to the existence of pre-existing schemas. That is, by activating the corresponding pre-existing schemas, the acquisition of newly input information is enhanced (Diekelmann & Born, 2010; Klinzing et al., 2019; Tse et al., 2007). Specifically, Diekelmann and Born (2010) explained that there are two distinctive memory stores: the temporary store and the long-term store. As the names represent, the temporary store learns information quickly but can retain memory temporarily whilst the long-term store requires longer time to learn it but can form a long-lasting memory. During the consolidation processes, the memory in the temporary store is re-activated. The important feature is, this re-activation also triggers re-activation of the corresponding memory in the long-term store. This means that when there is a similar memory traces exist in the long-term memory, this representation is also re-activated during the consolidation processes. Through this re-activation of the corresponding areas in the long-term store, newly-acquired memories are gradually copied within the long-term store. Diekelmann and Born described the function of these re-activation processes by using the self-explanatory metaphor; re-activation of memories in the temporary store is an internal "trainer" for the long-term store, which helps the long-term store to adapt newly-acquired memories into the pre-existing representations in this memory store. In this way, labile memory can be transformed into a more durable form and its processes relate to the availability of pre-existing memory representations in the long-term memory. Note that the temporary store and the long-term store are sometimes represented as the hippocampus and the neocortex since these areas are responsible to store each type of memory (e.g. Klinzing et al., 2019; Tse et al., 2007).

Whilst memory studies (e.g. Lewis & Durrant, 2011; Tuckey & Brewer, 2003) demonstrated the ability of schema-congruent memories to survive the passage of time and bolstered the important function of pre-existing schemas in consolidation processes, the amount of research that has examined the influence of sleep consolidation processes on the robustness of colour-word congruency has been insufficient. Despite that, congruency with schemas could suggest a possible scenario for how such a memory transformation process might affect ink colour effects on the AMST when such effects are measured by memory accuracy. In the present research it is possible to intuitively speculate that the colour-word schema exists with congruent colours (e.g. red ink and the word *strawberry*) rather than with incongruent colours (e.g. blue ink). With this assumption, it is still unknow whether the memory of congruent colour-word presentations can be consolidated better via sleep compared to incongruent colour-word presentations. That is, on the same day, the ink colour effects with colour-related words may be too weak to produce effects on memory performance whilst such effects might become more robust via sleep consolidation processes and generate effects on memory tests.

The present experiment examined whether introducing the consolidation processes via sleep would modify the previously-reported no ink colour effects with colour-related words in the present thesis. Thus, this experiment aimed to assess whether the robustness of ink colour effects during the AMST that measured as recognition memory accuracy changes with the experience of sleep consolidation processes since the memory of colour-related words with congruent ink colours may survive the passage of time whereas those with incongruent ink colours might disappear. To achieve this aim, the participants carried out the study over 2-days. Day 1 followed the same procedure as Experiment 2; the participants were asked to remember the eight pairs of colour-related words and numbers on the AMST for the later memory tests. These memories were tested on a cued-recall test, which asked them to type the corresponding numbers when the words appeared on screen one at a time. At the same time on the next day (i.e. the 24-hour interval), they experienced Day 2 that consisted of two parts: cued-recall test part and recognition test part. In the former part the same procedure was the same as Day 1 except the participants studied a new set of the word-number pairs (e.g. *sapphire-4*). Upon the completion of the cued-recall test part, they experienced the recognition test (i.e. asked them to judge whether they had seen the presented word-number pairs before or not) that examined memory accuracy of all word-number pairs (i.e. the word-number pairs from Day 1 and Day 2). The effects of the consolidation processes on task-irrelevant colour-word associations in the AMST were analysed with memory accuracy scores (i.e. d’ values) on this recognition test. The reason for selecting 24 hours as the interval to experience the consolidation processes was to ascertain that the participants had enough hours of sleep for such processes to occur. Based on Diekelmann and Born’s (2010) review article, which introduced a series of studies observing memory enhancement via a wide range of sleep hours (e.g. 6-minute short nap, nap for 1 or 2 hours and 8-hour nocturnal sleep; Korman et al., 2007; Lahl, Wispel, Willigens, & Pietrowsky, 2008; Mednick, Nakayama, & Stickgold, 2003; Nishida & Walker, 2007; Tucker et al., 2006), longer hours of sleep seem to produce larger facilitation effect on memory performance (Gais, Plihal, Wagner, & Born, 2000; Stickgold, Whidbee, Schirmer, Patel, & Hobson, 2000; Walker et al., 2003). From these, it is possible to assume that if the experiment was conducted between the morning to the early evening on Day 1, it would increase the chances of having some hours of nocturnal sleep (e.g. 8 hours), which may induce the consolidation processes. If the interval was set longer than 24 hours, there would be a risk of forgetting the materials that were studied on Day 1 and showed no recognition memory accuracy differences in all three colour congruency conditions (i.e. congruent, incongruent and neutral). Since it is still unclear whether memories of the word-number pairs that were presented on Day 1 could survive the 24-hour interval or not yet, the present experiment selected this duration as the starting point. The hypotheses in the present experiment were that, to items that were presented on Day 1, the congruent condition might show the higher recognition memory test scores than the neutral condition whilst the incongruent condition would reveal poorer memory accuracy than the neutral condition.

Method

Participants

Ninety students of University of Kent at Canterbury were recruited as the participants (i.e. 30 participants in each of the three colour congruency conditions). Since the word-number pairings in all colour congruency conditions were completely different from the previous experiments in the present thesis, the prior participation experience for Experiment 1–5.B was not considered. In the congruent condition there were four male participants (*M* = 18.75 years old; *SD* = 0.96) and 26 female participants (*M* = 18.88 years old; *SD* = 0.91). The incongruent condition consisted of five male participants (*M* = 22.20 years old; *SD* = 2.95) and 25 female participants (*M* = 19.32 years old; *SD* = 1.11) whilst in the neutral condition there were three male participants (*M* = 19.67 years old; *SD* = 0.58) and 27 female participants (*M* = 19.19 years old; *SD* = 1.36). This group of 90 participants consisted of 56 native English speakers (Congruent: *n* = 18; Incongruent: *n* = 19; Neutral: *n* = 19), 27 non-native English speakers (Congruent: *n* = 10; Incongruent: *n* = 9; Neutral: *n* = 8) and seven “English plus other language(s)” speakers (Congruent: *n* = 2; Incongruent: *n* = 2; Neutral: *n* = 3).

Design

The present experiment was a 3 (Colour congruency condition; congruent, incongruent, neutral) x 2 (Sleep consolidation; occurred, not occurred) factorial design with two independent variables: Colour congruency (a between-participant factor) and sleep consolidation (a within-participant factor). The main dependent variables were the total numbers of correct and incorrect responses on a recognition test. Scores on the cued-recall tests and reaction times on these two types of memory tests were also recorded. The assignment of colour congruency condition to each participant was counterbalanced based on Williams Latin Square Design (Williams, 1949).

Materials

The materials were the same as Experiment 4 except the following points. Day 1 employed the same eight colour-related words as Experiment 4 whilst Day 2 used a new set of eight colour-related words (see Appendix H and I). These additional colour-related words were selected based on three studies conducted by Huettig and Altmann (2011), Joseph and Proffitt (1996) and Tanaka and Presnell (1999). For the blue-related word, the word *sapphire* was selected from Huettig and Altmann’s (2011) study. Note that they also employed the word *jeans* as a blue-related word; however, due to the wider colour variety of jeans, *sapphire* was selected for the present experiment. The word *aubergine* was chosen for the purple-related word based on Joseph and Proffitt's experiments (1996), which used a line drawing of an aubergine as a purple-related concept. Thus, the word *aubergine* was employed in the present experiment. The other six additional colour-related words were obtained from Tanaka and Presnell's study (1999) for their high percentages of word-colour agreement. The additional colour-related words for Day 2 are listed here: *sapphire*, *horse*, *broccoli*, *basketball*, *aubergine*, *tomato*, *baseball* and *banana*. The average Zipf value of these eight Day 2 colour-related words was 3.95 and the average length of letters was 7.50 letters.

The same eight digits (i.e. from *1* to *8*) were used as the previous experiments in the present thesis whilst half of them was presented in the Roman numerals. This new numeral system was employed in order to reduce the confusion between the learning materials. If the participants' memory scores were extremely low in all colour congruency conditions due to the confusion between the to-be-studied items, ink colour congruency would not be able to have an influential impact on memory performance; rather, poor memory accuracy would have occurred due to the similarity between the materials. In order to prevent this situation, half of the numbers was written in the different numeral system (i.e. the Roman numerals). Across both days, the blue-related words, brown-related words, green-related words and orange-related words were paired with the Arabic numerals whilst the purple-related words, red-related words, white-related words and yellow-related words were paired with the Roman numerals. When the four numbers were presented in the Arabic numerals on Day 1 (e.g. *1*, *4*, *5*, *7*), these numbers were written in the Roman numerals on Day 2 (e.g. *I*, *IV*, *V*, *VII*) and vice versa.

The incongruent ink colour patterns were different between Day 1 and Day 2. For instance, one participant saw the *sky*-*6* pair with the number *6* was written in yellow ink on Day 1 and studied the *sapphire*-*4* pair with the number *4* was presented in brown ink on Day 2 (see Appendix H and I). This use of different ink colours as incongruent ink colours was performed to prevent the extreme poor memory accuracy due to the confusion between the learning materials.

The present experiment created the new pairings of words and numbers (see Appendix H and I). In terms of the numeral system assignment, the same four digits in the same numeral system were used to create the correct pairs and the incorrect pairs. For instance, if the study list on Day 1 included the *sky-1* pair, the *carrot-5* pair, the *grape-III* pair and the *strawberry-VIII* pair in the congruent condition, the recognition test consisted of these pairs as the correct pairs and different pairings as the incorrect pairs (e.g. *sky-5* and *grape-VIII*). On the recognition test, Day 1 items refer to the correct word-number pairs that were retrieved from Day 1 and the incorrect word-number pairs that consisted of different pairings with the words and numbers from Day 1. Day 2 items indicate these two types of pairs with the words and numbers from Day 2.

At the end of Day 2, the similar one-trial ink colour congruency check questionnaire was inserted in order to measure the participants' awareness of the ink colour difference as Experiment 5-B did. Since in this experiment each participant experienced only one type of colour congruency, less response keys (i.e. four response options) were used (1 = *matched*, 2 = *not matched*, 3 = *black*, 4 = *I don't remember*) in the present experiment compared to Experiment 5-B.

Procedure

The present experiment consisted of two parts: the cued-recall test part (Day 1 and 2) and the recognition test part (Day 2 only) (see Figure 3.3.1). In the 48-trial number-typing practice sections, the digits in half of the trials were presented in the Arabic numerals whilst the rest was written in the Roman numerals. After this practice section, Day 1 followed the similar procedure as Experiment 2 (i.e. three study-test sessions, cued-recall tests). They were also instructed to remember these word-number pairs for the subsequent memory test on the next day. Between Day 1 and Day 2, there was a 24-hour interval (e.g. if a participant started the experiment at 9 a.m., this person began the Day 2 part at 9 a.m. on the next day).

On Day 2, before starting the experiment, the experimenters asked the participants' hours of sleep to ascertain that they experienced sleep. Upon the completion of the same 48-trial number-typing practice section, the participants followed the same cued-recall study-test procedure as Day 1 with the new set of word-number pairs. After they finished the third session, the same 16-trial tone-comparison task was inserted as the previous experiments in the present thesis. This tone-comparison task was followed by the recognition test part, which procedure was identical to Experiment 3 and 4. In this recognition test part the participants did not need to specify whether the presented pairs were from Day 1 or Day 2. Since this recognition test included the word-number pairs on both days, it consisted of 32 test trials [(8 correct pairs + 8 incorrect pairs) x 2 days]. Upon the completion of the recognition test, the participants underwent the identical one-trial ink colour congruency check questionnaire as Experiment 5-B with less response options.

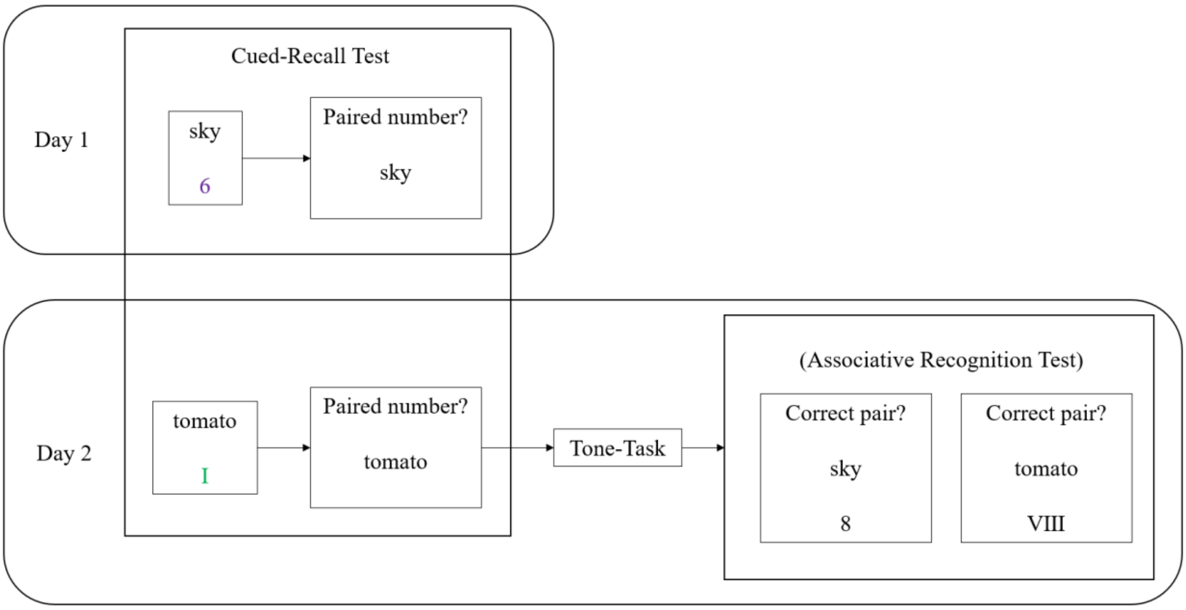


Figure 3.3.1. The general procedure of Experiment 6 (incongruent condition). Cued-recall test sections on both days consisted of three study-test sessions.

Consequently, Day 1 consisted of 24 cued-recall test trials (8 word-number pairs x 3 sessions) and Day 2 included 24 cued-recall test trials, 16 tone-comparison trials, 32 recognition test trials and one ink colour congruency check questionnaire trial. Day 1 took approximately 10 minutes to complete and Day 2 took approximately 20 minutes to finish.

Results

Mean proportion correct responses on the cued-recall tests, hit (Hit) rates, false alarm (FA) rates, d' values, c values on the recognition tests and mean reaction times on both memory tests are summarised in Table 3.3. On both days, none of the participants was timed out during the memory tests. On the cued-recall tests, all participants' responses were made in longer than 200 ms on the first day whilst four participants made responses in less than 200 ms on the second day. However, one of them made such responses only in two trials and the other three made those only in one trial out of 24 test trials. Thus, it would be possible to assume that these signified their appropriate attention to the task. On the recognition test, there were two participants who responded in less than 200 ms; however, each of them made such a quick reaction only in one trial out of 32 trials. Since this could also indicate the participants' appropriate attention to the task, all of their responses on both tests were included in the following analyses. To confirm whether participant in each colour congruent condition had the identical amount of sleep, the one-way ANOVA with Colour congruency as a between-participant factor was performed with hours of sleep. Since the main effect of Colour congruency was not significant [$F(2, 87) = 1.43$, $p = .244$, partial-$\eta^2 = .03$], this indicated that hours of sleep was identical in all colour congruency conditions [Congruent; $M = 7.13$, $SD = 1.63$; Incongruent; $M = 7.25$, $SD = 1.49$; Neutral; $M = 6.57$, $SD = 1.87$]. Therefore, it was confirmed that the following results occurred not due to the different amount of sleep in each colour congruency condition.

Table 3.3

Mean Proportion Correct Responses and Reaction Times (in ms) on the Cued-Recall Tests and Hit Rates, False Alarm Rates, d' Values, c Values and Reaction Times (in ms) on the Recognition Test on Day 1 and Day 2 in Each Colour Congruency Condition (Experiment 6)

		Cued-Recall						Recognition				
		Mean Proportion Correct			Reaction Time			Hit Rate	FA Rate	d'	c	Reaction Time
		1st	2nd	3rd	1st	2nd	3rd					
Condition	Day	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)
Congruent	Day 1	.47 (.25)	.75 (.20)	.91 (.12)	2,631 (1,039)	2,206 (698)	1,817 (546)	.78 (.14)	.33 (.21)	1.42 (0.93)	-0.17 (0.38)	1,970 (530)
	Day 2	.50 (.18)	.72 (.24)	.80 (.20)	2,239 (549)	2,089 (663)	1,872 (583)	.81 (.19)	.18 (.18)	2.11 (0.99)	0.03 (0.43)	2,003 (509)
Incongruent	Day 1	.44 (.21)	.73 (.18)	.84 (.18)	2,471 (760)	2,204 (780)	2,085 (811)	.75 (.21)	.40 (.25)	1.12 (1.29)	-0.24 (0.40)	2,049 (614)
	Day 2	.56 (.22)	.75 (.17)	.88 (.12)	2,101 (554)	2,000 (733)	1,834 (664)	.77 (.24)	.27 (.27)	1.68 (1.55)	-0.07 (0.39)	1,984 (612)
Neutral	Day 1	.53 (.25)	.71 (.26)	.83 (.18)	2,626 (813)	2,230 (637)	1,786 (621)	.72 (.17)	.40 (.27)	1.05 (1.16)	-0.18 (0.40)	2,008 (680)
	Day 2	.51 (.22)	.68 (.23)	.78 (.22)	2,233 (600)	2,315 (923)	2,043 (693)	.72 (.28)	.23 (.20)	1.62 (1.42)	0.09 (0.44)	1,978 (465)

Note. The values on the recognition test represent performance when responding to Day 1 items and Day 2 items.

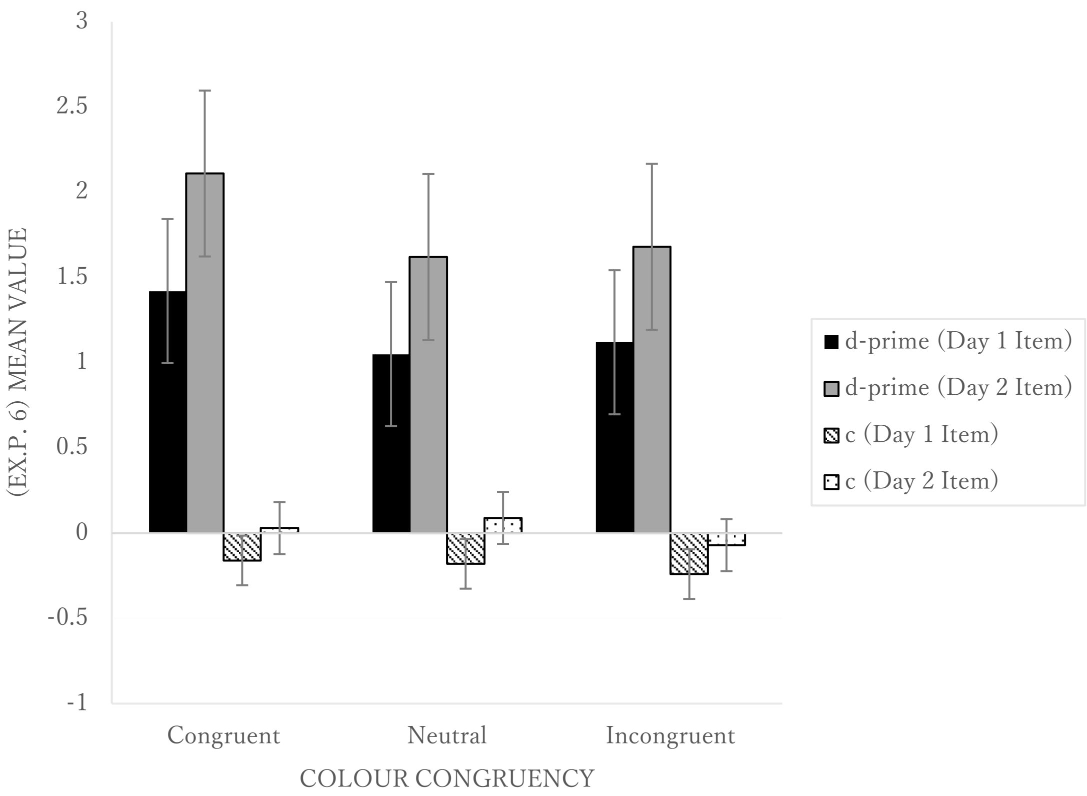


Figure 3.3.2. The graph shows the mean d’ values and the mean c values on Day 1 and 2 in each colour congruency condition (Experiment 6). Error bars represent the 95% confidence intervals (Masson & Loftus, 2003).

Analysis of Accuracy

The 3 (Colour Congruency) x 2 (Day) factorial ANOVA with Colour congruency as a between-participant factor and Day as a within-participant factor with d’ values was conducted (see Table 3.3 and Figure 3.3.2). The analysis revealed the significant main effect of Day, $F(1, 87) = 20.02$, $p < .001$, partial-$\eta^2 = .19$ whilst the interaction effect of Colour congruency x Day was not significant, $F(2, 87) = 0.09$, $p = .914$, partial-$\eta^2 = .002$. With Bonferroni correction, the significant main effect of Day indicated that the participants’ d’ values were significantly higher when responding to Day 2 items ($M = 1.81$, $SD = 1.34$) compared to Day 1 items ($M = 1.20$, $SD = 1.16$). However, Figure 3.3.2 seems to have facilitation effects between the congruent condition and the neutral condition with Day 1 and Day 2 items, the two one-way ANOVA for each day items was conducted. The analyses again revealed no significant main effect of Colour congruency to both types of items [Day 1 items: $F(2, 87) = 0.85$, $p = .431$, partial-$\eta^2 = .02$; Day 2 items: $F(2, 87) = 1.17$, $p = .315$, partial-$\eta^2 = .03$]. Planned contrasts confirmed no facilitation occurred with Day 1 items [$F(1, 87) = 1.51$, $p = .223$, partial-$\eta^2 = .02$] and with Day 2 items [$F(1, 87) = 1.96$, $p = .165$, partial-$\eta^2 = .02$]. When the analyses included only the congruent and incongruent condition, the 2 (Colour congruency) x 2 (Day) factorial ANOVA showed the same pattern of the results; the significant main effect of Day [$F(1, 58) = 14.70$, $p < .001$, partial-$\eta^2 = .20$] and the insignificant main effect of Colour congruency [$F(1, 58) = 1.83$, $p = .181$, partial-$\eta^2 = .03$] and interaction effect [$F(1, 58) = 0.16$, $p = .695$, partial-$\eta^2 = .003$],

For response bias, the same 3 x 2 factorial ANOVA was conducted with c values (which demonstrate correct/incorrect response tendency). The analysis also demonstrated the same pattern; the significant main effect of Day [$F(1, 87) = 10.08$, $p = .002$, partial-$\eta^2 = .10$] and the insignificant interaction effect of Colour congruency x Day [$F(2, 87) = 0.22$, p

= .806, partial-η^2 = .01]. The main effect of Day indicated that the participants' response bias values to Day 2 items (*M* = 0.02, *SD* = 0.42) was significantly higher than to Day 1 items (*M* = -0.20, *SD* = 0.40). This result demonstrated that the participants had the close-to-neutral response tendency when responding to Day 2 items whilst they showed the higher tendency to answer "correct" to Day 1 items. This could be due to the more superior memory accuracy with Day 2 items compared to Day 1 items. Since the participants had seen all words and numbers that consisted of the correct pairs and the incorrect pairs, it was reasonable that they showed the tendency to answer "correct" (i.e. the feeling of "I have seen this before") to all items. However, due to the accurate memory to Day 2 items, it was assumed that this response tendency was negated when responding to this type of the word-number pairs; thus, the participants revealed the close-to-neutral response tendency to Day 2 items whilst they tended to show more "correct" responses when responding to Day 1 items. To examine whether there was difference in response bias between each colour congruency condition, the two one-way ANOVA were performed but revealed no main effects of Colour congruency to the items from Day 1 [$F(2, 87) = 0.25$, $p = .783$, partial-η^2 = .01] and from Day 2 [$F(2, 87) = 1.11$, $p = .334$, partial-η^2 = .03]. This confirmed that colour congruency had no impact on response bias on the recognition test in the present experiment. To sum up, whilst the participants' response tendency to answer "correct" or "incorrect" varied depending on the type of stimuli (i.e. from Day 1 or Day 2), the response bias occurred identically between the three colour congruency conditions.

Analysis of Accuracy with Performance on the Cued-Recall Tests

For performance on the cue-recall tests (see Table 3.3), one might question whether the participants' memory accuracy of Day 1 items affected the later recognition memory tests. That is, if they did not remember Day 1 materials accurately on the first day, performance on the recognition test might have reflected the effects of Day 2 items only;

thus, the insignificant ink colour effect and higher memory accuracy to Day 2 items were observed. To test this, the same 3 x 2 factorial ANOVA was conducted with those who scored 100% on the final session [Congruent: $n = 17$, Incongruent: $n = 12$, Neutral: $n = 11$]. However, the analysis still revealed the same pattern: the significant main effect of Day [$F(1, 37) = 5.91$, $p = .020$, partial-$\eta^2 = .14$] and the insignificant interaction effect of Colour congruency x Day [$F(2, 37) = 1.71$, $p = .195$, partial-$\eta^2 = .09$]. This could confirm that although the participants did not remember Day 1 items perfectly, it would not have a huge impact on performance on the following recognition test.

Analysis of Awareness

Whether the participants' awareness of ink colour congruency affects memory performance was examined with d' values and correctness on the ink colour congruency check questionnaire. This was tested with the 3 (Colour congruency) x 2 (Day) x 2 (Ink Awareness) factorial ANOVA with Ink awareness as a between-participant factor whilst 54 participants correctly recognised all types of colour congruency. Since the analysis demonstrated that the interaction effect of Colour Congruency x Day x Ink Awareness was not significant [$F(2, 84) = 0.45$, $p = .637$, partial-$\eta^2 = .01$], this could indicate that the participants' awareness of ink colour congruency did not modify the interaction effect of colour congruency and sleep consolidation processes on memory accuracy.

Discussion

The present experiment examined whether introducing the consolidation processes via sleep would modify the previously-reported no ink colour effects with the colour-related words. To find an answer for this question, the aim of Experiment 6 was to study whether presenting the word-number pairs with congruent ink colours (e.g. *sky-1* with blue ink) could survive the passage of time and reveal more superior memory accuracy on the subsequent recognition test since these were expected to fit the pre-existing schema; therefore, would be consolidated whilst presenting with incongruent ink colours may not survive due to the lack of pre-existing schemas and consolidation. The hypothesis was that, when responding to Day 1 items, the congruent condition would reveal the higher recognition memory test scores than the neutral condition whilst the incongruent condition would reveal the lower one than the neutral condition. To test this, the participants experienced the two-day experiment. On Day 1, the participants were asked to remember the eight pairs of colour-related words and numbers (e.g. *sky-1*) on the AMST. Day 2 followed the same procedure with the new set of word-number pairs (e.g. *sapphire-8*) but included the recognition test for all word-number pairs at the end. The numbers of correct responses on this recognition test were analysed. The analyses indicated two main findings; the participants showed (1) more superior memory accuracy to Day 2 items than Day 1 items and (2) identical memory accuracy in all three colour conditions to both Day 1 and Day 2 items. Thus, the hypothesis was not supported.

The first outcome could be explained by the perspective of the recency effect (A. D. Baddeley & Hitch, 1993). The participants' higher scores to Day 2 items than to Day 1 items might have occurred due to the phenomenon that the recently presented items (e.g. the last word) are recalled better; this circumstance is called the recency effect (A. D. Baddeley & Hitch, 1993). In the literature Baddeley and Hitch asserted that the recency effect stems from the mechanism that is identical to the priming effect (i.e. a phenomenon that occurs when the previously-presented items modify the subsequent task performance) and implicit learning (i.e. acquiring information in the absence of consciousness; Reber, 1989, see also Richardson-Klavehn & Bjork, 1988). This expected connection between the recently-presented materials and unconscious information processes could offer a potential explanation for the reported result in the present experiment (i.e. Day 2 items were recognised more accurately than Day 1 pairs). Specifically, it would be possible to presume that the priming and implicit learning occurred to the strategy of remembering word-number pairs. On Day 1, the participants were asked to remember the word-number pairs for the later memory tests and the same procedure was performed in the first part of Day 2. This repetition of the same task across the two days may have signalled the occurrence of priming and implicit learning on Day 2; that is, the participants had unconsciously developed the learning strategy to remember the presented word-number pairs efficiently by the experience on Day 1 and this skill was transferred to Day 2. Simply, when the participants were attempting to remember the materials, the task expertise level was low on Day 1 whereas it reached at an advanced stage by the time they studied the items on Day 2. Consequently, the word-number pairs that were presented at a later date were remembered more accurately.

No ink colour effects and the higher memory test scores to Day 2 items would suggest the inappropriateness of sleep as a tool to compensate the weak colour-word relationships with colour-related words. Evidence from the present experiment indicated that sleep seems to induce forgetting rather than consolidation for the learning materials on the AMST. Whilst one might question whether the shorter or longer interval (e.g. 12 hours, 48 hours or one week) reveals any ink colour effects, the literature (e.g. Diekelmann & Born, 2010) has been reporting the wide variety of evidence in terms of the duration and timing of sleep. Such a wide range of manipulating sleep would signal the difficulty of keeping exploring the effects of sleep consolidation processes with the AMST at a deeper level without the confirmation that it is unable to reveal immediate ink colour effects from colour-related words. Therefore, the following experiments returned to the one-day format and investigated the immediate effects of task-irrelevant ink colours with the AMST. To explore such effects, the tentative recency-effect account signals the possibility of involving priming with the AMST. That is, since the previous exposure to the materials affected the level of memory accuracy and the strength of semantic links between colour concepts and word meanings seems to have an influential impact in the AMST, these raise a question; whether pre-activation of semantic connections between colours and colour-related words can strengthen their relationships. Thus, Experiment 7 used the priming approach to analyse this question.

Experiment 7

The present experiment examined whether previous-exposure of information could be a potential factor that the existing model of cognitive load theory needs to include. As introduced in Chapter 1, previous-encounter of paired-associate learning stimuli (i.e. word-word pairs) can modify the subsequent task performance (Spieler & Balota, 1996), which could signal the involvement of such effects in the field of paired-associate learning. Additionally, Experient 6 suggested the involvement of priming when ink colour effects with colour-related words on the AMST are measured by memory tests. Priming represents a phenomenon when the previous exposure to stimuli modifies the subsequent processes of the same materials or related ones (Tulving, Schacter, & Stark, 1982). Generally, there are four types of priming: conceptual priming, perceptual priming, semantic priming and associative priming. Conceptual priming occurs when there is a conceptual relationship between the previously-presented stimuli and the presented materials (Huntjens et al., 2002). The example would be when reading the priming word *ROBIN* changes a response to the subsequent word *EAGLE* (Lucas, 2000). Since these two concepts share the same features, these are conceptually related (Lucas, 2000); thus, this type of priming is named conceptual priming. Perceptual priming occurs when the subsequent responses are modified due to changes in physical appearance that are necessary to form a shape of stimuli (e.g. line elements, print typography of words but not colours) between the previously-encountered items and the present stimuli (Wiggs & Martin, 1998). Semantic priming occurs when there is a meaning relationship between the previously-exposed items and the present items (Lucas, 2000). Specifically, if one concept is necessary to constitute another, it could be possible to determine that these are related semantically. For instance, calcium is an essential chemical element that constitutes milk; thus, although these two concepts belong to different categories (e.g. calcium might belong to a chemical component category whilst milk could be a member of drink), the meanings of these are related. Therefore, if the word *CALCIUM* is previously presented and this previous encounter modifies participants' response to the word *MILK* on the subsequent task, this case would be considered as semantic priming (for an overview see Lucas, 2000). Associative priming occurs due to word association without categorical congruency between items (e.g. *needle-thread*). That is, if there is an instrumental association between the priming materials and the subsequently-presented items (e.g. *needle-thread*), this is the case of associative priming (for an overview see Lucas, 2000). Whilst the specific criterion of these types have been equivocal due to overlaps between their concepts (for an overview see Lucas, 2000), pre-activation of colour-word relationships would be categorised into semantic priming since colours constitute the objects that the words refer to (e.g. *red-strawberry*). Therefore, the present experiment employed semantic priming to modify the robustness of such an association.

To produce priming effects in the Stroop-related tasks, there are two important factors to consider: the number of pre-activated dimensions and the relevancy between pre-activated processes and the present task. The former factor was suggested by McClain (1983) who examined the types of priming words and its effect on the robustness of the subsequent Stroop task. As McClain found the largest Stroop interference effect in the colour name/secondary word condition (e.g. the priming word was *red* and the Stroop word was *blood*) and the same secondary word condition (e.g. the priming word was *apple* and the Stroop word was *apple*) whilst the weakest effect occurred in same non-colour-word condition (e.g. the priming word was *nail* and the Stroop word was *nail*), the researcher concluded that the Stroop interference effect becomes larger when more dimensions are pre-activated (i.e. the word *red* pre-activates word and colour dimensions). The latter factor was retrieved from MacLeod's (1996) study, which examined the priming effects on a colour-naming task and a word-reading task. The participants were asked to remember the presented priming words in a neutral ink colour (e.g. *clan*) and after this, they underwent a colour-naming task and a word-reading task whilst the to-be-processed words included the priming words (i.e. *clan*). Note that the order of these tasks was counterbalanced. For colour naming, the robustness of the Stroop effect was unchanged regardless of the type of the Stroop words (i.e. previously-seen or new). However, the researcher found a facilitation effect from the priming words on word reading; previous exposure to the materials accelerated the speed of reading words on the subsequent task. Based on these results, MacLeod (1996) concluded that priming is process-specific. That is, when participants see the priming words, their task is to remember the words; thus, they are required to "read" the priming words. When the following task requires the same process (i.e. word reading), priming effects are likely to occur but if not (i.e. colour naming), the prior exposure to the materials has little impact on the subsequent task. In this way, MacLeod highlighted the importance of the relevancy between pre-activated processes and the present task to produce priming effects.

Whilst McClain (1983) and MacLeod's (1996) studies illustrated the essential elements of priming, Levin and Tzelgov's (2016) Experiment 4 indicated that priming has an influential impact on inducing effects from colour-related words on the Stroop task. Since the researchers failed to observe the Stroop effect with colour-related words in their previous experiments, the aim of the fourth experiment was to examine whether priming of colour-word associations by showing coloured pictures of colour-related objects (e.g. a picture of a red tomato) assists colour-related words to produce the Stroop effect on colour naming. Note that these colour-related concepts were used as the Stroop words in the later colour-naming task. This priming succeeded in intensifying the robustness of colour-word associations with colour-related words. When the stimulus categories were separated by blocks, previous exposure to the pictures of colour-related words induced the Stroop interference effect (i.e. slower response speed with incongruent ink colours than with neutral ink colours) from this type of colour words; priming overturned the previously-reported no ink colour effects. Therefore, Levin and Tzelgov concluded that in the blocked format, colour-related words require the pre-enhancement of colour-word associations to produce interference effects on a later task.at first glance, their use of coloured pictures (i.e. seeing pictures) seemed to pre-activate the different processes during the Stroop task (i.e. colour naming) whilst they observed priming effects on the robustness of colour-word relationships. However, pre-activation of verbal information during priming could be the same triggered process as when the participants read the Stroop words on the later colour-naming task. According to the dual-code theory (Paivio, 1971, 1986, 2007, as cited in Kousta, Vigliocco, Vinson, Andrews, & Del Campo, 2011; Paivio, 1991, 2006, 2013; also see McBride & Dosher, 2002), processing pictorial stimuli tends to involve their appearance information and generated labels whilst processing words also triggers verbal information. Therefore, when the participants in Levin and Tzelgov's (2016) study saw the pictures during priming, it would be possible to assume that they generated labels of the pictures (e.g. "tomato"), which information was also activated during the Stroop task. Thus, the same processes (i.e. activating verbal information of the stimuli) seemed to be activated across the priming stage and the main task.

The conclusions from these three studies (Levin & Tzelgov, 2016; MacLeod, 1996; McClain, 1983) suggested a possibility that priming might induce task-irrelevant ink colour effects from colour-related words in the AMST. Consideration of the key factors to produce priming effects might suggest one question; when colour-related words are previously presented (i.e. activating two dimensions) and colour-word links are pre-activated, whether it would help colour-related words to produce ink colour effects when such effects on the AMST are measured by memory tests. Whilst the literature (Levin & Tzelgov, 2016; MacLeod, 1996; McClain, 1983) has reported effects on a speed dimension (although Levin and Tzelgov reported performance on the recognition tests, the purpose of this analysis was to confirm that the participants had remembered the pictures accurately), Chapter 2 in the present thesis illustrated that such effects can be applicable to an accuracy aspect; therefore, the present experiment applied the findings from the priming literature to examine memory accuracy with the AMST.

The present experiment examined whether introducing priming induces task-irrelevant ink colour effects from colour-related words when memory tests measures such effect on the AMST. The aim of Experiment 7 was to examine whether pre-activation of colour-word associations via semantic priming induces task-irrelevant ink colour effects with colour-related words when such effects are examined by recognition memory test. To achieve this aim, the two types of priming were inserted before each study section: a semantic-priming task and a lexical-priming task. In a former case, the pairs of colour-related words (which were different from the to-be-remembered materials) and coloured patches were presented. The participants' task was to judge whether these were related (i.e. the pair of the word *silver* and silver-coloured patch required a "correct" response) or not. The reason why the priming materials were different from the main studied items was since the focus of this research was memory accuracy. If the same colour-related words (i.e. *strawberry*) were presented during priming and the main task, this could create the situation that the participants had an established memory of the words before they entered the study section; this prior experience only with the words might function as extra practices. To prevent such inequal amount of practice between the colour words and paired numbers, the present experiment employed different colour-related words between priming and the main study section. In a lexical-priming task the priming words were presented with a single alphabet and the participants were asked to judge whether the words included the paired alphabet (i.e. *silver*-*i*) or not (e.g. *fork*-*a*). The hypothesis was that only after the semantic-priming task, the incongruent condition would reveal poorer memory accuracy than the neutral condition. The expectation was made only for interference since only this type of effects have consistently emerged in the previous experiments in the present thesis. Therefore, such an effect would be expected to happen at a higher chance than facilitation.

Method

Participants

Sixty students of University of Kent at Canterbury were recruited as the participants. Since the current word type (i.e. colour-related words) was different from Experiment 5-A and 5-B (i.e. emotional words), the prior participation experience for these two experiments was not considered whilst all participants had no prior experience of participating in Experiment 1-4 and 6. There were eight male participants (*M* = 18.00 years old; *SD* = 0.00) and 52 female participants (*M* = 18.83 years old; *SD* = 1.17). Thirty participants were assigned to the semantic-priming task (Male: *n* = 1; Female: *n* = 29) and the rest of them was given the lexical-priming task (Male: *n* = 7; Female: *n* = 23). The mean ages of each priming group were *M* = 18.63 years old with *SD* = 1.10 in the semantic-priming group and *M* = 18.80 years old with *SD* = 1.16 in the lexical-priming group. This group of 60 participants consisted of 51 native English speakers (Semantic: *n* = 26; Lexical: *n* = 25), eight non-native English speakers (Semantic: *n* = 4; Lexical: *n* = 4) and one “English plus other language(s)” speakers (Lexical: *n* = 1).

Design

The present experiment was a 3 (Colour congruency condition; congruent, incongruent, neutral) x 2 (Priming; semantic, lexical) factorial design with two independent variables: Colour congruency (a within-participant factor) and Priming (a between-participant factor).

Materials

The majority of the materials were the same as Experiment 2 and 4 (see Appendix B) whilst the same ink colour congruency check questionnaire was used as in Experiment 5-B. The difference from Experiment 1-6 was that the present experiment included the priming materials. To pre-activate colour-word links, the semantic priming task demonstrated pairs of colour-related words and coloured patches in the form of the AMST. To keep the word type consistent throughout the experiment, the word type was limited to colour-related words. Since this priming task was inserted in all three study-test sessions, the goal of stimulus selection was to have 24 colour-related words (8 pairs x 3 sessions). The researcher attempted to retrieve the colour words that satisfy with the following stimulus selection standards as many as possible; the words that (a) can be both colour names and colour-related words, (b) have the identical letter length as Chapter 2 studies, (c) have the highest or close-to-highest word frequency values and (d) can trigger a specific image colour. For example, the word *daisy* can trigger a colour of white (Digital Synopsis, n.d.) and of yellow (Huettig & Altmann, 2011). Since the semantic priming task asked participants to judge whether colour words and coloured patches are correctly associated, such an activation of various colours would confuse the participants. To prevent this, if the words activate several image colours, those were excluded. With using the websites ("Colour," 2020; W3Schools, n.d.), the database (Brysbaert et al., 2014; van Heuven et al., 2014) and the literature (Huettig & Altmann, 2011) to examine these standards, 19 words were obtained from the website that show lists of colour words (Digital Synopsis, n.d.). Then five additional words were retrieved from the website (i.e. *chocolate*; W3Schools, n.d.) and the literature (i.e. *salmon*; Huettig & Altmann, 2011, *pig* and *fork*; Tanaka & Presnell, 1999 and *rat*; Joseph & Proffitt, 1996). In total, 24 colour-related words were selected (see Appendix J) and are listed here in the alphabetical order: *amber*, *amethyst*, *aquamarine*, *bronze*, *brunette*, *chocolate*, *emerald*, *fork*, *gold*, *honey*, *ivory*, *lavender*, *lime*, *linen*, *marigold*, *olive*, *orchid*, *pearl*, *pig*, *rat*, *ruby*, *salmon*, *silver* and *turquoise*.

In the semantic-priming tasks these colour-related words were presented with coloured patches. The majority of the ink colours of these patches were the exact colours that the words refer to (e.g. the colour patch paired with the word *silver* was presented in silver ink); that is, no modification in their RGB codes. However, RGB codes for some of the ink colours were modified in order to increase their distinctiveness. For instance, the colours marigold and honey appeared to be similar whilst marigold is in the list of orange and honey is in the yellow-related colour word list on the website (Digital Synopsis, n.d.). To make these colours more orange and yellow, RGB codes were changed by the researcher. In terms of the pink-related (i.e. *pig*) and silver-related words (i.e. *fork* and *rat*), these were not colour names. Therefore, three ink colours were arbitrarily employed from the website (Digital Synopsis, n.d.) and HTML colours (W3Schools, n.d.): magenta, navy and pink. All ink colours that were used in the coloured patches are summarised in the alphabetical order [RGB codes in square brackets]: amber [230,172,0], amethyst [153,102,204], aquamarine [127, 255, 212], bronze [185,114,45], brunette [77,38,0], chocolate [210, 105, 30], emerald [80,200,120], gold [255, 215, 0], honey [169,131,7], ivory [255,255,215], lavender [188,188,242], lime [0, 255, 0], linen [246,227,209], magenta [255, 0, 255], marigold [238,182,80], navy [0, 0, 128], olive [128, 128, 0], orchid [153,50,204], pearl [234,224,200], pink [255, 192, 203], ruby [244,17,95], salmon [250, 128, 114], silver [192, 192, 192] and turquoise [0,206,209].

To create three lists of priming materials, the experimenter randomly selected one colour-related words from each colour category (e.g. red-related) via the website (Randomness and Integrity Services Ltd, n.d.). Using the same website, these words were randomly paired with the coloured patches and assigned to one of two word-item pair types (i.e. correct pair and incorrect pair) for the semantic-priming tasks. For the lexical-priming tasks, these colour-related words were also randomly assigned to one word-item pair type and randomly paired with one alphabet, which was included in the words (e.g. the letter *n* in the word *aquamarine*), for the correct word-item pair type. In the incorrect word-item pair type the words were paired with other random alphabets that the words did not include (e.g. the letter *a* with the word *fork*).

Procedure

The procedure in the present experiment was similar to Experiment 4 except the participants experienced the priming tasks before entering the study section and were asked to complete the ink colour congruency check questionnaire at the end as in Experiment 5-B. Specifically, after they finished the practice section, the screen informed that the participants would see colour-related words (e.g. *sapphire*, *rose*, *tomato*, *frog*). Then they were assigned to one of the priming groups (i.e. semantic or lexical) and saw the task instruction with the examples. The screen instructed them to judge whether the word and coloured patch were related (in the semantic-priming group) or the word included the alphabet (in the lexical-priming group). The instruction screen also told the participants to respond as quickly and accurately as possible. Note that the number of priming trials were not provided to the participants. After this instruction, the first priming task began with no time limits. The priming words were presented in the same font, size, position and black ink as the main to-be-studied words whilst the colour patches and alphabets were shown in the same size and position as the to-be-studied numbers. In both priming tasks half materials required “correct” responses and half required “incorrect” responses with the same assigned keys as in the main recognition tests. Upon completion of the first priming task, the participants followed the same procedure as Experiment 4. When they finished the 16-trial tone-comparison task, the priming task instruction was shown again and the second priming task was started. In total, there were 24 priming trials (8 pairs x 3 sessions), 48 recognition test trials (16 pairs x 3 colour congruency conditions), 32 tone-comparison trials (16 trials x 2 sessions) and one ink colour congruency check trial. The present experiment took approximately 15 minutes to complete.

Results

The following analyses focused only on the recognition test. The participants' performance on the priming tasks was not analysed since the aim of these tasks was to induce pre-activation of colour-word relationships from the semantic-priming tasks and compare its effect with non-semantic-priming group (i.e. the lexical-priming group). Thus, as long as the participants considered semantic connections in the semantic-priming group or lexical components in the lexical-priming group, their task performance during the priming tasks was not the main interest of this study. In the present experiment none of the participants was timed out and all reaction times on the recognition test were longer than 200 ms, which indicated their appropriate attention to the task. Thus, all responses were included in the following analyses.

Mean hit (Hit) rates, false alarm (FA) rates, d' values, c values and reaction times on the recognition tests after the priming tasks are summarised in Table 3.4.

Table 3.4

Hit Rates, False Alarm Rates, d' Values, c Values and Reaction Times (in ms) in Each Colour Congruency Condition with the Sub-Groups of the Priming Task Type (Experiment 7)

Condition	Priming Task Type	Hit Rate	FA Rate	d'	c	Reaction Time
		M (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)
Congruent	Semantic	.69 (.15)	.39 (.17)	0.84 (0.76)	-0.12 (0.31)	2,122 (1,011)
	Lexical	.72 (.18)	.34 (.18)	1.15 (0.96)	-0.10 (0.29)	2,049 (655)
Incongruent	Semantic	.64 (.14)	.31 (.15)	0.92 (0.73)	0.07 (0.25)	2,159 (957)

	Lexical	.69 (.16)	.33 (.16)	1.05 (0.86)	-0.03 (0.27)	2,005 (580)
Neutral	Semantic	.67 (.16)	.33 (.17)	0.98 (0.90)	0.002 (0.24)	1,849 (633)
	Lexical	.72 (.14)	.34 (.18)	1.10 (0.83)	-0.07 (0.25)	1,909 (493)

Note. The values on the recognition test represent performance after the semantic-priming tasks or the lexical-priming tasks.

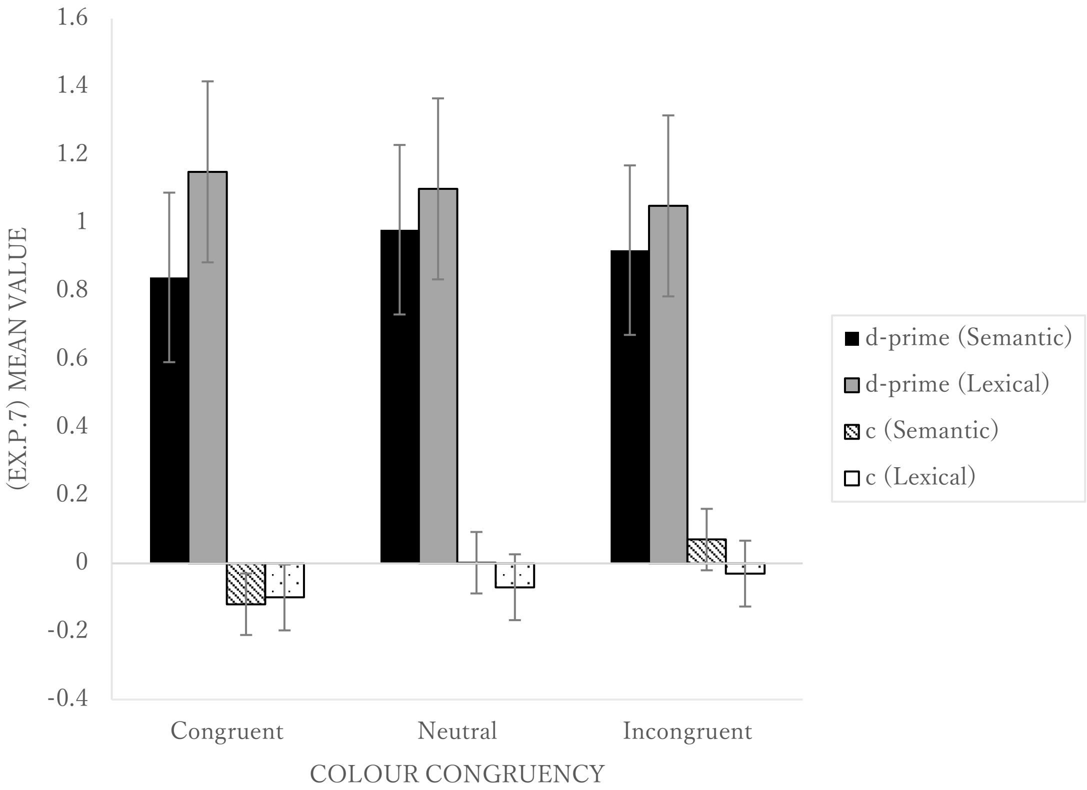


Figure 3.4. The graph shows the mean d' values and the mean c values after the semantic-priming tasks or the lexical-priming tasks in each colour congruency condition (Experiment 7). Error bars represent the 95% confidence intervals (Masson & Loftus, 2003).

Analysis of Accuracy

As in Experiment 6, d' values (which represent recognition memory accuracy) were analysed with the 3 (Colour Congruency) x 2 (Priming) factorial ANOVA with Colour congruency as a within-participant factor and Priming as a between-participant factor (see Table 3.4 and Figure 3.4). The analysis with d' values revealed that there was no significant main effect of Colour congruency, $F(2, 116) = 0.10$, $p = .905$, partial-$\eta^2 = .002$, and no significant interaction effect of Colour congruency x Priming, $F(2, 116) = 0.33$, $p = .723$, partial-$\eta^2 = .01$. However, Figure 3.4 seems to have facilitation and interference effects after the lexical-priming tasks and also suggested a possibility of observing interference after the semantic-priming tasks. Therefore, the two one-way ANOVA with d' values for each priming group was conducted. The analyses again revealed no significant main effect of Colour congruency after both types of priming tasks [Semantic: $F(2, 58) = 0.29$, $p = .747$, partial-$\eta^2 = .01$; Lexical: $F(2, 58) = 0.14$, $p = .869$, partial-$\eta^2 = .01$]. When comparing the congruent and incongruent conditions only, the 2 (Colour congruency) x 2 (Priming) factorial ANOVA replicated the results; no significant main effect of Colour congruency [$F(1, 58) = 0.01$, $p = .925$, partial-$\eta^2 = .00$] and no interaction effect of Colour congruency x Priming [$F(1, 58) = 0.46$, $p = .499$, partial-$\eta^2 = .01$].

The same 3 (Colour Congruency) x 2 (Priming) factorial ANOVA was conducted with c values (which demonstrate correct/incorrect response tendency) and demonstrated the interesting pattern; the significant main effect of Colour congruency [$F(2, 116) = 3.78$, $p = .026$, partial-$\eta^2 = .06$] but the insignificant interaction effect of Colour congruency x Priming [$F(2, 116) = 1.01$, $p = .367$, partial-$\eta^2 = .02$]. Paired sample *t*-tests (two-tailed) indicated that the participants' response bias values in the congruent condition ($M = -0.11$, $SD = 0.30$) was significantly lower than in the incongruent condition ($M = 0.02$, $SD = 0.26$), $t(59) = -2.62$, $p = .011$, $d = 0.34$. The difference between the congruent and neutral conditions ($M = -0.03$, $SD = 0.25$) was not significant [$t(59) = -1.54$, $p = .129$, $d = 0.20$] and between the neutral and incongruent conditions [$t(59) = -1.27$, $p = .211$, $d = 0.16$]. These results demonstrated that the participants had the close-to-neutral response tendency when responding to the word-number pairs, which had been presented with the incongruent ink colours, whilst they showed the higher tendency to answer "correct" to the items, which had been presented with the congruent ink colours. Such a more robust "correct" response tendency in the congruent condition could be explained by the words that had been presented during the priming tasks. In both priming groups the participants saw the colour-related words such as *aquamarine* or *ivory*. Since exposure to the colour words appears to activate colour-word associations (Hazan-Liran & Miller, 2017), this activation could be expected to trigger the semantic connections between typical image colours and colour words. Due to this pre-activated colour-word relationships with congruent image colours, the participants might have accidentally felt the stronger sense of familiarity to the word-number pairs, which had been presented with the congruent ink colours (Feenan & Snodgrass, 1990).

Analysis of Awareness

To analyse the effects of ink colour awareness on task performance, the 3 (Colour Congruency) x 2 (Priming) x 2 (Ink Awareness) factorial ANOVA was conducted with Ink awareness as a between-participant factor whilst 24 participants (Semantic: n = 9 and Lexical: n = 15) correctly recognised all types of colour congruency. This revealed the insignificant interaction effect of Colour congruency x Priming x Ink awareness, $F(2, 112)$ = 1.38, p = .256, partial-η^2 = .02. This could indicate that the effects of ink colours on recognition memory accuracy were independent from the participants' awareness of colour congruency.

Discussion

The present experiment examined whether priming can overturn the previously-reported no ink colour effects with colour-related words on the AMST. Thus, it aimed to study whether pre-activation of colour-word associations leads colour-related words to produce effects when such effects on the AMST are measured by recognition memory test. The hypothesis was that after semantic priming, the participants would reveal interference due to pre-strengthened colour-word associations. The methodology was identical to Experiment 4 with two priming tasks; the semantic- or lexical-priming task. These were performed before each study section and memory performance on the recognition tests were analysed. The analyses with d' values replicated the results in the previous experiments; memory accuracy was identical regardless of priming type and ink colour congruency. Whilst the results with response bias might have signalled the successful priming effect on pre-activating colour-word links, such enhancement seemed to be not robust enough to affect d' values. Therefore, the present analyses did not support the hypothesis.

Whilst data in the present experiment may suggest the successful pre-activation of colour-word association, the use of new colour-related words as priming materials might have contributed to no ink colour effects. In Levin and Tzelgov's (2016) study they used the same colour-related concepts as priming materials (e.g. a picture of tomato) and the Stroop words (e.g. the word *TOMATO*). Since the same colour-concept association was activated at priming and at colour naming, this seems to fulfil both key factors to produce priming effects in the Stroop-related research; pre-activate the larger number of dimension (McClain, 1983) and increase the relevancy between pre-activated processes and the present process (MacLeod, 1996). Thus, this difference could indicate that the robust pre-activation of colour-word associations is required to produce effects with colour-related words when memory tests are combined with the AMST. To produce effects, the possible solution would be using the same colour-related words during priming tasks and the main task. If performance on the main task is measured by response speed, this modification would be appropriate. However, when measuring memory accuracy, there would be a potential risk that the use of the same materials between priming tasks and the main memory task may induce the unequal amount of exposure to to-be-remembered stimuli. Alternatively, presenting colour-related concepts in a different modality (e.g. pictorial format) could be another way to strengthen colour-word relationships. Since information that is presented in a pictorial format tends to show greater tolerance to memory decay than in a verbal format (Peeck, 1974), this may have contributed to Levin and Tzelgov's (2016) study. That is, the participants had to perform the colour-naming task (e.g. naming the ink colour of the word *TOMATO*) whilst they retained the memories of the coloured pictures of the concepts (e.g. a coloured picture of tomato). Thus, presenting colour-related concepts in a pictorial format might be the next approach to examine the potential factor that the existing model of cognitive load theory needs to include.

In summary, using priming replicated no ink colour effects from the colour-related words on recognition memory accuracy as the previous experiments in the present thesis have revealed. Despite this, it could indicate that the AMST might require the robust activation of colour-word links when the words are colour-related words and effects are measured by memory tests. Building upon these, the next experiment employed pictures of colour-related objects. In this way, the next one examined how using different modality of materials modifies the previously-reported no ink colour effects with colour-related words.

Experiment 8

The present experiment explored whether the stimulus presentation format contributes to inducing effects from colour-related words when memory tests measure such effects on the AMST. Chapter 1 introduced that changing the stimulus presentation format (e.g. presenting materials as a picture instead of written words) is one way to enhance paired-associate learning (Paivio & Yarmey, 1966). Additionally, the previous experiment used written words as priming materials and demonstrated no ink colour effects whilst Levin and Tzelgov's (2016) successfully induced ink colour effects from colour-related concepts by using pictures as priming materials to strengthen colour-concept links. From the contradicting results between Experiment 7 in the present thesis and the literature (Levin & Tzelgov, 2016), this may suggest a possibility that the stimulus modality might affect the robustness of colour-word associations when memory is studied with the AMST.

Compared to written words, the literature has reported more robust effects on memory performance with pictures. Such effects were named picture superiority effect (i.e. presenting information in a pictorial format enhances recall memory accuracy compared to a verbal format; Nelson et al., 1976). At encoding, this advantage of pictures could be explained by the dual-code theory (Paivio, 2006), which proposes dual memory storage for pictorial information in imaginal and verbal formats. Pictures also demonstrate a beneficial effect on retention. For instance, Peeck (1974) studied the effects of illustrations on learning and revealed stronger resistance to memory decay with pictorial contents than verbal contents. Such a longer memory retention might have occurred due to the availability of pre-existing schemas. As Bentin and Moscovitch (1988) reported, memory is enhanced when there is an available pre-existing schema for the newly-acquired information. Simply, more familiar stimuli create stronger and more accessible memory traces. Thus, in Peeck's (1974) finding of longer retention with pictorial presentation might be because information is stored in memory as a pictorial or visual format (i.e. having picture-concept schemas) rather than verbal presentation; therefore, learning materials as pictures matches with the pre-existing schema, which might have resulted in longer retention with a pictorial format. Whilst research has demonstrated superior memory performance with pictures, one might question whether object colour information is included in pictorial memories and if so, how such attached colour information affects processes in the Stroop-related studies. The literature (Naor-Raz, Tarr, & Kersten, 2003) has provided evidence for these questions; colour information is co-activated when visual information is processed, which produces the Stroop effect with pictorial stimuli. When an object is robustly associated with a typical colour (e.g. a banana and yellow), seeing the shape of this object automatically activates its typical colour. Due to this automatic activation of colour-shape relationships, it interferes with colour naming when this object is presented in incongruent colours (e.g. a picture of a purple banana). Additionally, such effects of stored colour knowledge can occur with uncoloured monochrome pictures (Joseph & Proffitt, 1996); the memory of "a banana is yellow" affects task performance when participants see the coloured or uncoloured pictures of a banana.

These findings might offer a way to induce effects with colour-related words when such effects on the AMST are measured by memory tests. The literature would suggest two key findings; (1) presenting information in a pictorial format enhances memory accuracy than in a verbal format and (2) seeing coloured or monochrome pictures of colour-related concepts trigger the typical image colour of these objects. When combining these, it may suggest a question to examine; whether presenting colour-related concepts in a pictorial format activates the stronger memory of colour-concept relationships than in a verbal format.

Therefore, the present experiment explored whether using pictures induces any ink colour effects from colour-related concepts when memory tests measure such effects on the AMST. The aim of Experiment 8 was to examine whether studying the picture-number pairs shows any ink colour effects from colour-related concepts when effects on the AMST are measured as recognition memory accuracy whilst studying the word-number pairs replicates the previously-reported no ink colour effects. To achieve this aim, the participants were assigned to one of two groups: the picture format group or the word format group. The former group was asked to remember pairs of monochrome pictures and numbers on the AMST (e.g. saw the monochrome picture of the pig and the number *4* with *4* was printed in pink, yellow or black ink). The word group studied the same pairs but in the word-number format (e.g. saw the pairs of the word *pig* and the number *4*). After study, these memories were examined on the recognition test and response accuracy was analysed. As interference has stably observed in Chapter 2, the hypothesis was particularly made to this effect; when studying pairs of pictures and numbers, the incongruent condition would show lower recognition test scores than the neutral condition.

Method

Participants

Sixty students of University of Kent at Canterbury were recruited as the participants. As in Experiment 7, their prior participation experience for Experiment 5-A and 5-B was not considered whilst they had no prior experience of participating in Experiment 1-4, 6 and 7. There were 10 male participants (M = 22.50 years old; SD = 11.43) and 50 female participants (M = 19.52 years old; SD = 3.71). Thirty participants were in the picture format group (Male: n = 6; Female: n = 24) and the rest of them in the word format group (Male: n = 4; Female: n = 26). The mean ages of each stimulus presentation format were M = 19.83 years old with SD = 4.62 in the picture format and M = 20.20 years old with SD = 6.70 in the word format. This group of 60 participants consisted of 50 native English speakers (Picture: n

= 25; Word: *n* = 25), nine non-native English speakers (Picture: *n* = 4; Word: *n* = 5) and one "English plus other language(s)" speakers (Picture: *n* = 1).

Design

The design of the present experiment was identical to Experiment 7 except the following point. The present experiment was a 3 (Colour congruency condition; congruent, incongruent, neutral) x 2 (Stimulus; picture, word) factorial design with two independent variables: Colour congruency (a within-participant factor) and Stimulus (a between-participant factor).

Materials

The majority of the materials were identical to Experiment 2, 4, 6 and 7 except the following point. The difference from these previous experiments was the inclusion of pictorial materials. Note that the picture format group and the word format group studied the exact same concept-number pairs; difference was only in the stimulus presentation format. To show the pictures clear enough for the participants to identify, the pictures, words and numbers in this experiment were presented in a larger size than Experiment 1-7. The size of all materials were increased equally since presenting each in different sizes might affect how the participants process these items (e.g. Rammsayer & Verner, 2015). The pictures and words were shown in the width and length of 0.4 points at around the centre of the screen (i.e. 0.2 points above the central point) in Psychopy (Peirce et al., 2019). The paired numbers were presented in the same letter size but slightly below the central point of the screen (i.e. 0.3 points below in Psychopy). The new format of materials (i.e. the picture format) was used in the practice section and the study-test sections. The criteria of selecting these pictures was whether Snodgrass and Vanderwart's (1980) list of 260 pictures included these since their pictures have been widely used as stimuli (e.g. Alvarez & Cavanagh, 2004; Rasamimanana, Barbaroux, Colé, & Besson, 2020).

For the practice section, the words *soup* and *song* (which were used in the previous experiments) were replaced with new words and pictures since these words were not included in Snodgrass and Vanderwart's (1980) picture list. The two new words were retrieved from the two original studies (Hazan-Liran & Miller, 2017; P. Miller et al., 2018) with the selection criteria of being included in Snodgrass and Vanderwart's list (see Appendix K). Whilst the word *basket* was not included in both of the original studies (Hazan-Liran & Miller, 2017; P. Miller et al., 2018), the picture and verbal label of it were employed due to its conceptual similarity to the word *bucket*, which was used in the two studies (Hazan-Liran & Miller, 2017; P. Miller et al., 2018).

For the to-be-studied materials, since some of the previously-used colour-related words were not included in Snodgrass and Vanderwart's picture list and it was difficult to precisely present these in the monochrome pictures (e.g. *sky*), four colour-related concepts (i.e. *sky*, *mud*, *grass*, *cream*) were replaced with new concepts: *pig*, *horse*, *lettuce* and *swan*, respectively (see Appendix L and M). Since it was difficult to accurately present the sky and other blue-related concepts (e.g. sapphire, jeans; Huettig & Altmann, 2011) in monochrome pictures, this blue-related colour concept was removed and the pink-related concept (i.e. *pig*) was added. The associated ink colour was pink ink [255, 192, 203 in RGB code]. The words *pig*, *horse* and *lettuce* were obtained from the literature (Tanaka & Presnell, 1999) for their high percentages of word-colour agreement. However, due to the low percentage of white-related word in this literature, the word *swan* was retrieved from a new study (Naor-Raz et al., 2003) for its existence in Snodgrass and Vanderwart's picture list and identifiability of the concepts (i.e. whether the picture is easy to be understood), which was determined arbitrarily. The angles and directions of all pictures were the same as the original ones (e.g. the face of a pig pointed to the left and the leaf of a carrot appeared on top). Since the main focus of the present research is difference in memory accuracy due to colour congruency, as long as these patterns were kept congruent across all colour congruency conditions, it was expected to have no impact on the analyses of memory accuracy; therefore, these elements followed the original presentation.

Procedure

The procedure in the present experiment was similar to Experiment 7 except there were no priming tasks and the participants in the picture format group studied the pairs of the pictures and numbers (e.g. a monochrome picture of a pig and the number *4*). In total, there were 48 recognition test trials (16 pairs x 3 colour congruency conditions), 32 tone-comparison trials (16 trials x 2 sessions) and one ink colour congruency check trial. The present experiment took approximately 15 minutes to complete.

Results

In the present experiment none of the participants was timed out and all reaction times on the recognition tests were longer than 200 ms, which indicated their appropriate attention to the task. Therefore, all responses were included in the following analyses. Mean hit (Hit) rates, false alarm (FA) rates, d’ values, c values and reaction times on the recognition tests with each stimulus presentation format are summarised in Table 3.5.

Table 3.5

Hit Rates, False Alarm Rates, d’ Values, c Values and Reaction Times (in ms) in Each Colour Congruency Condition with the Sub-Groups of the Stimulus Presentation Format (Experiment 8)

Condition	Stimulus Presentation Format	Hit Rate	FA Rate	d’	c	Reaction Time
		M (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)
Congruent	Picture	.67 (.15)	.33 (.18)	1.04 (0.87)	0.02 (0.32)	2,115 (908)
	Word	.69 (.15)	.39 (.16)	0.88 (0.80)	-0.12 (0.27)	2,106 (749)
Incongruent	Picture	.71 (.15)	.26 (.15)	1.35 (0.91)	0.06 (0.25)	1,967 (530)
	Word	.71 (.16)	.34 (.17)	1.11 (0.86)	-0.10 (0.30)	1,997 (856)

Neutral	Picture	.70 (.17)	.30 (.19)	1.22 (1.07)	-0.004 (0.26)	2,050 (698)
	Word	.70 (.17)	.39 (.16)	0.94 (0.93)	-0.16 (0.23)	1,868 (738)

Note. The values represent performance when the stimuli were presented in the picture-number format or the word-number format.

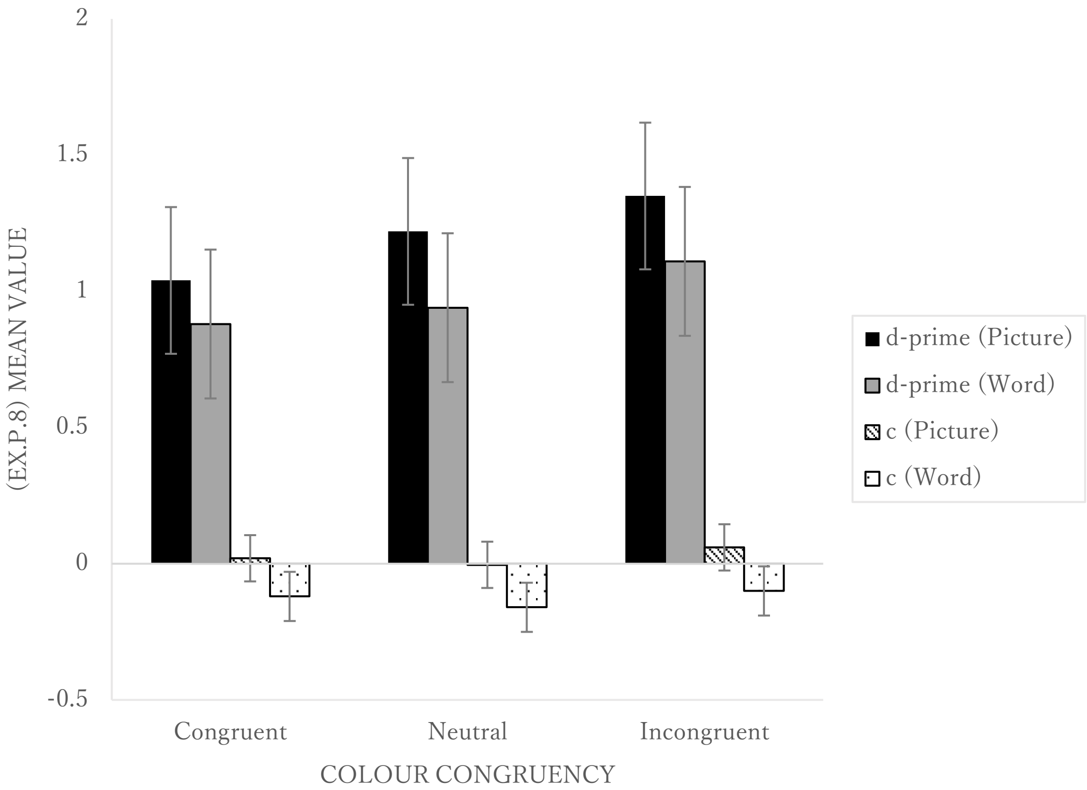


Figure 3.5. The graph shows the mean d' values and the mean c values when the stimuli were presented in the picture-*number* format or the *word-number* format in each colour congruency condition (Experiment 8). Error bars represent the 95% confidence intervals (Masson & Loftus, 2003).

Analysis of Accuracy

The d' value difference between the congruent and incongruent conditions with two stimulus types was examined. Since the 2 (Colour congruency) x 3 (Stimulus) factorial ANOVA demonstrated the significant main effect of Colour congruency [$F(1, 58) = 4.58$, $p = .037$, partial-$\eta^2 = .07$] whilst the interaction effect was not significant [$F(1, 58) = 0.11$, $p = .741$, partial-$\eta^2 = .002$], the following analyses compared memory performance in each colour congruency condition with the neutral condition. As in Experiment 7, d' values (which represent recognition memory accuracy) were analysed with the 3 (Colour Congruency) x 2

(Stimulus) factorial ANOVA with Colour congruency as a within-participant factor and Stimulus as a between-participant factor (see Table 3.5 and Figure 3.5). The analysis with d' values revealed that there was no significant main effect of Colour congruency, $F(2, 116) = 1.98$, $p = .142$, partial-$\eta^2 = .03$, and no significant interaction effect of Colour congruency x Stimulus, $F(2, 116) = 0.10$, $p = .905$, partial-$\eta^2 = .002$. Since Figure 3.5 seems to have a significant difference between the two stimulus presentation formats, the mean d' values were re-calculated for each participant by removing colour congruency factor. The difference between the two stimulus presentation format groups was not significant, $F(1, 58) = 1.66$, $p = .203$, partial-$\eta^2 = .03$.

The same 3 (Colour Congruency) x 2 (Stimulus) factorial ANOVA was conducted with c values (which demonstrate correct/incorrect response tendency). The analysis demonstrated no significant main effect of Colour congruency [$F(2, 116) = 1.98$, $p = .142$, partial-$\eta^2 = .03$] and the insignificant interaction effect of Colour congruency x Stimulus [$F(2, 116) = 0.10$, $p = .905$, partial-$\eta^2 = .002$]. However, the analysis revealed a significant difference across the two stimulus presentation formats, $F(1, 58) = 9.30$, $p = .004$, partial-$\eta^2 = .14$. Removal of the colour congruency factor showed that the picture format group ($M = 0.04$, $SD = 0.24$) had the significantly higher response tendency values than the word format group ($M = -0.12$, $SD = 0.18$). This result demonstrated that the participants had the close-to-neutral response tendency when studying the pairs of pictures and numbers whilst they showed the higher tendency to answer “correct” to the pairs of words and numbers. Whilst such a difference was significant, this could be due to the higher frequency of seeing the pairs of words and numbers than the picture-number pairs. As learners become older, the stimulus presentation of learning materials tends to transform from pictures to words (e.g. picture book to textbooks). Since the participants were university psychology students, it was possible to assume that the majority of information is presented in the word format rather than the pictorial presentation. This could imply that when they study knowledge that is related to numbers (e.g. statistics), such numerical information is more likely to be presented with words than pictures; seeing numbers with words might be more common than with pictures. Because of this, the overall level of familiarity was high for the word-number pairs, which resulted in the lower c values and contributed to a significant difference in response tendency between the two stimulus presentation formats (Feenan & Snodgrass, 1990).

Analysis of Awareness

To examine whether the participants' awareness of ink colour congruency affect d' values, the 3 (Colour Congruency) x 2 (Stimulus) x 2 (Ink Awareness) factorial ANOVA was conducted whilst 37 participants (Picture: n = 17; Word: n = 20) correctly recognised all types of ink colour congruency. This revealed that the interaction effect of Colour congruency x Stimulus x Ink awareness was not significant [$F(2, 112) = 1.07$, $p = .348$, partial-$\eta^2 = .02$], which indicated that whether the participants realised ink colour congruency during study or not had no effects on the interaction effect of Colour congruency x Stimulus.

Discussion

This experiment explored the question whether the use of pictures instead of words produces ink colour effects from colour-related concepts when such effects on the AMST are measured by memory tests. It aimed to examine whether effects of task-irrelevant ink colours would be observed from the picture-number pairs on the recognition tests. The hypothesis was that the incongruent condition would reveal poorer memory accuracy than the neutral condition only in the picture format group. To test this hypothesis, the participants were asked to remember either the pairs of the monochrome picture and the coloured number (e.g. the picture of a pig and the number *4* with *4* was presented in pink, yellow or black ink) or the word-number pairs (e.g. the *pig-4* pair with the word *pig* was written in black ink but *4* was in pink, yellow or black ink) and these memories were tested on the recognition tests. The analyses replicated no ink colour effects regardless of the stimulus presentation format; the hypothesis was not supported.

The contradicting results between the present experiment and the literature (Levin & Tzelgov, 2016; Naor-Raz et al., 2003) might suggest that the AMST requires the robust activation of colour-concept relationships such as by including colours in pictures in order to produce effects on memory performance. The present experiment failed to induce the effects of colour-concept links with the pictorial stimuli whereas the studies that used the pictures as the Stroop stimuli (Naor-Raz et al., 2003) and priming materials (Levin & Tzelgov, 2016) succeeded to demonstrate such effects. Of particular note is that these researchers presented the coloured pictures. These successes with the coloured pictures might suggest the necessity of using this type of pictures in order to make colour-concept links strong enough to produce effects in the case of the AMST when studying memory. Evidence from the encoding specificity principle (Tulving & Thomson, 1973), which proposes memory enhancement by matching contexts of information between study and recall, could bolster this tentative account. Intuitively, it would be possible to assume that when people learn the typical colour of a strawberry, they are more likely to establish such a colour-concept link by seeing the actual object or the coloured picture of the item. Thus, although Huettig and Altmann's (2011) study demonstrated the ability of monochrome pictures to activate the stored colour knowledge, due to congruency of contexts between study and recall, correctly coloured pictures (e.g. the picture of a strawberry is coloured in red) are expected to activate more robust colour-concept associations than monochrome pictures. Thus, when measuring ink colour effects on the AMST by memory tests, the AMST might require such a robust activation to produce the effects. This encoding specificity account might not only offer an explanation for the contradicting results between the present experiment and the literature

(Levin & Tzelgov, 2016; Naor-Raz et al., 2003) but also suggest difficulty of using pictorial stimuli with the AMST (i.e. presenting the pictures in a neutral colour whilst the paired numbers are written in other colours) in memory research.

One might argue that shape similarity induced memory confusion in all colour congruency conditions in the picture stimulus format group as the literature (D. L. Nelson et al., 1976) demonstrated. Ink colour effects were not observed due to extremely poor memory accuracy in all colour congruency conditions. However, this was less likely to occur in the present experiment. If memory accuracy in all colour congruency conditions were too poor to show effects since all pictures looked similar in terms of shape, the pattern of memory accuracy was expected to be different from the word presentation format group, which was not affected by shape similarity between items. When analysing d’ values, the interaction effect of Colour congruency and Stimulus was not significant, which indicated that the pattern of the results was identical regardless of seeing pictures or words. Therefore, whilst exploring the properties of pictures on the AMST is still open to investigation, the present experiment would conclude that such an element is not a factor that contributes to no ink colour effects with colour-related concepts when the AMST is combined with memory tests.

Overall, Experiment 8 did not observe ink colour effects from colour-related concepts on memory accuracy with the AMST. Whilst further investigation of the visual properties of pictures remains open as a potential research topic, the present experiment may suggest inappropriateness of using pictures instead of colour words as a tool to strengthen colour-concept links. Therefore, it would be possible to conclude that the stimulus presentation format is less likely to be a factor that the existing model of cognitive load theory needs to include.

Chapter 3: Chapter Summary

Chapter 3 attempted to find an answer for the question; except semantic relatedness, what would be a potential factor that the existing model of cognitive load theory needs to include in order to examine the effects of extraneous cognitive load on the AMST in memory research. Thus, the present chapter examined in which circumstance colour-related words in the AMST might produce any effects on memory accuracy. Specifically, this was investigated from the perspective of word concreteness (Experiment 5-A and 5-B), sleep consolidation (Experiment 6), priming (Experiment 7) and the pictorial presentation format (Experiment 8). With these manipulations the ability of colour-related words/concepts to produce effects on recognition memory accuracy was tested using the AMST. The overall hypothesis was that through these manipulations, the incongruent condition would reveal poorer recognition memory accuracy than the neutral condition; that is, it was expected to find an interference effect. The majority of the procedure was identical to Chapter 2. The participants were asked to remember the presented item pairs and these memories were tested on an associative recognition test, which asked them to judge whether the pairing of the presented item pairs was correct or incorrect. With this procedure, Experiment 5-A used the pairs of colour names and emotional words (e.g. *white-happy*) and Experiment 5-B presented the pairs of negative emotion words and numbers (e.g. *fear-1*). Experiment 6 was a two-day study that the participants studied one set of word-number pairs (e.g. *sky-1*) on Day 1 and a new set (e.g. *sapphire-8*) on Day 2 and underwent a recognition test for both sets at the end of Day 2. The participants in Experiment 7 were given one of two priming tasks before each study section; a semantic priming task (i.e. asked them to judge whether the presented colour words and coloured patches were related) and a lexical priming task (i.e. asked them to judge whether the presented words included the simultaneously-presented letters). Those who were in Experiment 8 were divided into two groups; one group studied the pairs of monochrome pictures and numbers (e.g. the picture of a strawberry and the number *6*) and another saw the standard word-number pairs (e.g. *strawberry-6*). The overall result was that none of these manipulations induced ink colour effects from colour-related words/concepts. Therefore, this series of experiments did not support the hypothesis.

The overall finding in this chapter that colour-related words/concepts failed to produce robust effects seems to be in the same line with Levin and Tzelgov's (2016) study, which failed to observe the Stroop effect with colour-related words in three experiments. When one concept (e.g. FIRE) is associated with various concepts (e.g. RED, NATURE, HEAT), the strength of each semantic connection (e.g. FIRE-RED, FIRE-NATURE) becomes weaker as the number of co-activated concepts (i.e. semantic neighbours) increases (Buchanan, Westbury, & Burgess, 2001; Collins & Loftus, 1975); these weak colour-word associations seem to be a potential factor when colour-related words fail to produce ink colour effects on task performance. Whilst this could explain differences between colour names and colour-related words, it would be difficult to explain the contradicting results between the original studies (Hazan-Liran & Miller, 2017; P. Miller et al., 2018) and the present research when examining effects with colour-related words as the original studies reported effects whilst the present research has kept failing to do so. A possible factor to consider may be differences in the exposure duration during the study. That is, the participants in the present experiment were provided longer study times than Hazan-Liran and colleagues' experiments. For instance, according to Table 2 in Hazan-Liran and Miller's (2017) study, the participants in the congruent condition correctly completed 75.63 questions within 2 minutes. When simply dividing 120 seconds by the number of completion, it indicated that the participants spent on average 1.59 seconds to complete one response box. This clearly demonstrated that the participants in the original studies were exposed to word-number pairs for less time compared to those in the present research, which had provided 3 seconds to study the pairs. The literature (for an overview see Oberauer & Lewandowsky, 2011) has demonstrated that this difference in the duration of study affects the amount of cognitive load; with a longer available time for performing the task, the amount of cognitive load decreases. Combining the semantic account and the cognitive load account, it would suggest that the weak semantic relatedness and the extended duration of studying might contribute to the level of robustness of colour-word relationships with colour-related words with the AMST; therefore, the manipulations on word concreteness, memory consolidation and stimulus presentation formats failed to modify the robustness of colour-word associations in this chapter.

Therefore, evidence from the present chapter may imply that when enough study time is provided, the degree of overlap in meaning between learning materials and to-be-ignored stimuli might affect how accurately learners encode paired-associate information.

Whilst the current chapter illustrates a circumstance, which prevents the detrimental ink colour effects from colour-related words on memory performance, the previously-reported results in the present thesis were to explore what factors may contribute to the robustness of ink colour effects when such effects on the AMST are measured by memory tests. Therefore, it is still questionable what factor induces facilitation and interference on memory accuracy; how memory facilitation and interference occurred with colour names in Chapter 2. When considering effects of extraneous cognitive load on learning, the potential factor that the existing models of cognitive load theory are missing would be an emotional factor (for an overview see Plass & Kalyuga, 2019). Additionally, the study of the Stroop effect (e.g. Damen, Strick, Taris, & Aarts, 2018; Hatukai & Algom, 2017) has provided evidence for the relationship between the Stroop effect and emotions. Since Chapter 2 confirmed that effects of extraneous cognitive load can be also measured by memory performance and investigation of an emotional factor with the AMST has not been conducted yet even in the original studies (Hazan-Liran & Miller, 2017; P. Miller et al., 2018), it would be necessary to explore this factor with the AMST in order to make a conclusion for what factors the models of cognitive load theory need to include.

In summary, the present chapter demonstrates additional evidence for the applicability of the AMST by filtering the potential factors that the existing models of cognitive load theory need to consider. No ink colour effects with various manipulations may suggest the potential scenario that ink colour effects from colour-related words with the AMST are restricted to when pressure is largely placed upon response speed. As the present chapter seems to offer a tentative conclusion for the ink colour effects with colour-related words, the focus is now shifted to why colour names on the AMST revealed effects in Chapter 2. Thus, the next chapter explores what could be a key factor to produce memory performance difference with colour names when the AMST is combined with cued-recall tests.

Chapter 4: Emotional Influence on Colour Names on the AMST with Cued-Recall Tests

The present chapter mainly focuses on why presenting word-number pairs with incongruent ink colours impaired memory accuracy as this pattern of the effect was stably observed in Chapter 2. The primary account is derived from the perspective of traditional cognitive load theory, which proposes inhibition of task-irrelevant colour-word associations reduces the amount of available cognitive resources to learn word-number pairs on the AMST (Hazan-Liran & Miller, 2017; P. Miller et al., 2018). Whilst this account seems to be applicable to the present research, there would be another potential factor that might have contributed to such memory impairment: emotions. As an emotional factor has been reported to affect performance on the Stroop task (e.g. Hatukai & Algom, 2017), examining the influence of affective valence would be necessary to determine what factors the traditional cognitive load theory needs to include.

Whilst the Stroop literature has highlighted the importance of emotions, the traditional cognitive load theory seems to have insufficient consideration of emotional influences on learning efficacy (for an overview see Plass & Kalyuga, 2019). Other processing models of learning emphasise the importance of emotions (e.g. the Integrated Model of Cognitive-Affective Learning with Media; Plass & Kaplan, 2016) with the notion that humans have a tendency to avoid mentally or cognitively demanding work since mental or cognitive effort induces an aversive feeling (for an overview see Kool, McGuire, Rosen, & Botvinick, 2010). Therefore, Plass and Kalyuga (2019) argued that processing emotional information during learning would affect the amount of cognitive load, which determine the quality of learning. For example, in the case that the task requires a large amount of cognitive effort (e.g. solving the complex mathematical problem) and it induces an aversive feeling from learners, dealing with this negative emotion (e.g. suppressing the aversive feeling during calculation) is not relevant to the task whilst such inhibition requires additional cognitive resources. Thus, emotions during learning may contribute to the amount of extraneous cognitive load, which determine learning efficacy. Despite this assumption, different types of emotions have demonstrated opposite effects on the quality of learning. Specifically, learners' positive feelings tend to enhance learning whereas the negative emotions such as mental withdrawal or boredom are generally related to detrimental effects (for an overview see Pekrun, Goetz, Titz, & Perry, 2002). This different educational impact from types of emotions is also bolstered by Erk, Martin and Walter's (2005) study of brain activation; when seeing an emotionally-positive picture, the same neural circuits are activated at encoding and recognition stages whereas different brain areas seem to be involved during each process when learners see an emotionally-negative picture. Therefore, whilst it seems to be difficult to determine the direction of emotional influence (i.e. facilitation or interference) based on the cognitive load perspective and other learning literature, these findings clearly demonstrated the impacts of emotions on learning, which is related to the amount of cognitive load.

The review of the emotional impact on cognitive load and learning might suggest to include an emotional factor in the present research as it may offer a reason why interference occurs with the AMST. Since the AMST has the ability of producing semantic conflict between the meanings of ink colours printed in numbers and colour words, presenting word-number pairs with incongruent ink colours would generate a larger amount of cognitive demand that might signal the occurrence of negative emotions to this type of pairs (for a summary see Hatukai & Algom, 2017). Suppression of this task-irrelevant emotion required additional cognitive load; consequently, less available cognitive resources are left to remember the word-number pairs and memory accuracy in the incongruent condition was degraded. Since conflict seems to have an ability of inducing emotional reactions, this scenario would not be applied to the congruent and neutral conditions, which lacked semantic conflict. If this emotional account was the case, one might question whether manipulation of emotional information modifies performance on the AMST. This question might be tackled from the view of mood-congruent learning (V. E. Lewis & Williams, 1989; Plass & Kalyuga, 2019), which proposes that matching between learners' moods at study and affective valance of to-be-studied materials facilitates information-acquisition processes; feeling happy when studying would help them to learn materials that induce a happy feeling. Thus, it might be possible that increasing mood congruency between word-number pairs and learners' feelings might mitigate the detrimental impacts of task-irrelevant negative emotions during the task. Such a control of mood congruency can be tested by employing the similar techniques as the literature performed (Hatukai & Algom, 2017; van Steenbergen, Band, & Hommel, 2009).

Therefore, this chapter examined whether negative emotions induced by colour-word incongruency produce memory impairment when ink colour effects on the AMST are measured by memory tests. The aim of the present chapter was to clarify whether increasing unpleasantness of word-number pairs or learners' unpleasant feelings during study would mitigate the robustness of the interference effects with cued-recall tests. Since the methodology in the two original studies (Hazan-Liran & Miller, 2017; P. Miller et al., 2018) seems to involve the cued-recall processes (i.e. after they checked the target word-number pair in the example list, they returned to the copy-number test section in which required them to recall the associated numbers by cued colour names), this chapter returned to this close-to-original methodology. To achieve the aim, the affective valence of word-number pairs was manipulated by assigning "good" and "bad" impressions to each number (Experiment 9) whilst the learners' emotional states were controlled by inserting emotional stimuli (i.e. face icons in Experiment 10 and auditory stimuli in Experiment 11) during study. Note that since only colour names revealed effects, this chapter explored memory performance with only this type of colour word. Overall, it was hypothesised that, if negative emotions contribute to memory impairment with the AMST, increasing unpleasant impression of learning materials or learners’ unpleasant feeling would enhance the participants’ cued-recall memory accuracy in the incongruent condition due to mood congruency during study.

Experiment 9

The present experiment explored whether mood congruency between participants and affective valence of learning materials modifies performance on the AMST. Chapter 1 demonstrated the potential factors (i.e. word concreteness, consolidation, priming and stimulus presentation formats) that are involved in the field of paired-associate learning and the one (i.e. emotion) in the area of colour-word associations. However, Chapter 3 studies manipulated the former factors and revealed none of such manipulations modified performance on the AMST. This could suggest the possible scenario that emotional factors might be involved more when using the AMST. As the literature (Plass & Kalyuga, 2019) introduced in Chapter 4 Introduction showed the extension of this emotional influence to the field of cognitive load theory, the present experiment approached memory performance on the AMST from a perspective of emotions. Whilst the literature (V. E. Lewis & Williams, 1989; Plass & Kalyuga, 2019) has illustrated the facilitative effects of mood congruency on general learning-related tasks (e.g. remembering emotional words), such an emotional relationship between participants and stimuli can be observed in other cognitive tasks such as the Stroop task. The emotional effects in the Stroop task were named the Stroop Incongruity Effect by Hatukai and Algom (2017), who reported the mitigation of the Stroop interference effect by using “good” and “bad” responses on a colour-naming task.

Hatukai and Algom explored the question whether a third dimension of the Stroop stimuli affects performance on a colour-naming task. The traditional Stroop task involves only two dimensions: colours and words; therefore, it was unclear whether involving a third dimension modifies performance on this task. Whilst they found that responding to a third dimension (e.g. answering the vertical position of the incongruent Stroop stimuli) accelerated reaction times, the researchers suggested that this might involve avoidance of conflict that Stroop stimuli produce (Dreisbach & Fischer, 2012; Schouppe, De Houwer, Ridderinkhof, &

Notebaert, 2012; van Steenbergen et al., 2009). The incongruent Stroop stimuli (e.g. the word *RED* was printed in blue ink) induce an aversive feeling (Botvinick, Braver, Barch, Carter, & Cohen, 2001; Dreisbach & Fischer, 2012; Schouppe et al., 2012; van Steenbergen et al., 2009) and people respond to such a stimulus faster in order to minimise the exposure duration to this negative stimulus (Dignath & Eder, 2015; Goldfarb & Treisman, 2010; Kool et al., 2010; Schouppe, Ridderinkhof, Verguts, & Notebaert, 2014). Based on these assumptions, Hatukai and Algom argued that emotions (as a third dimension) determine performance on the traditional Stroop task and their argument was supported by their experiments, which used "good" and "bad" responses. For example, instead of naming ink colours of Stroop words, the participants in their Experiment 7 were asked to answer "good" to the congruent Stroop words (e.g. the word *RED* printed in red ink) and "bad" to the incongruent ones (e.g. the word *RED* printed in blue ink). In their Experiment 8 the "good" responses were required for the positive-valenced colours (i.e. red and green inks) and the "bad" responses were required for the negative-valenced colours (i.e. blue and brown inks) in one condition whilst these responses were reversed in another condition. Through these emotion labelling manipulations, the researchers found a significant emotion-response congruency. The reaction times became faster when saying "good" to the congruent Stroop words and positive-valenced colours than saying "bad" to these stimuli. For the "bad" responses, the response speeds were also accelerated when articulating "bad" to the incongruent Stroop words and negative-valenced colours (i.e. the colours predominantly presented in the incongruent stimuli) whilst articulating "good" to these stimuli made the responses slow. Additionally, their comparison of the congruent stimuli and incongruent ones demonstrated the reversal pattern of the Stroop effect; the responses were faster to the incongruent stimuli than the congruent ones when the responses involved the third dimension (i.e. good and bad). From these outcomes, Hatukai and Algom concluded that the incongruent Stroop stimuli are associated with a negative emotion and when responses involve the third dimension, this can modify the robustness of the Stroop interference effect.

Despite Hatukai and Algom found the faster reaction times by involving an emotional dimension, whether avoiding stimuli is possible or not determines the direction of emotional influences on the Stroop effect as they highlighted. That is, this facilitative effect from emotions might occur only when making a response allows participants to avoid or terminate the source of conflict whilst such an effect may not be observed when conflict is unavoidable (for a summary see Hatukai & Algom, 2017). This unavoidable conflict may contribute to memory impairment with the AMST as it is congruent with the associative memory literature, which described that emotional words require additional cognitive resources and less resources are left for binding the items pairs (Mather, 2007). If this unavoidable conflict contributes to memory impairment, memory performance is expected to be enhanced with mood-congruent since mood-congruent learning (V. E. Lewis & Williams, 1989; Plass & Kalyuga, 2019) proposes enhancement of task performance when learners' feelings match affective valence of to-be-remembered items. Additionally, the involvement of arousal in the Stroop task seems to be questionable. According to Russell's (2003) dimensional model of emotions, emotions consist of two dimensions: pleasantness and arousal. In the case of negative emotions, a feeling of "upset" and "sad" are both low on a pleasure scale whilst "upset" consists of high arousal and "sad" is a low-arousal emotion. Damen, Strick, Taris and Aarts (2018) demonstrated that seeing the incongruent Stroop words (e.g. the word *RED* in blue ink) triggered a low-pleasant feeling even when colour naming was not required. Their study indicated that simply seeing the incongruent Stroop stimuli make participants feel negative. However, their study examined only the Stroop effect on a pleasure dimension; the effect on an arousal dimension still seems to be unknown. Also, Hatukai and Algom's (2017) use of the good/bad responses might question an involvement of high arousal in the Stroop effect since, on the 9-point arousal rating scale (i.e. higher numbers indicate higher arousal levels; Warriner, Kuperman, & Brysbaert, 2013), the word *good* shows 3.66 and the word *bad* is 4.86 that indicate the close-to-neutral arousal levels. By contrast, the incongruent Stroop stimuli has been reported to induce an aversive feeling from participants (Botvinick et al., 2001; Dreisbach & Fischer, 2012; Hatukai & Algom, 2017; Schouppe et al., 2012; van Steenbergen et al., 2009) and an aversive feeling or situation is assumed to induce a certain level of arousal (Kim & Diamond, 2002). Combining these, it is still questionable whether a low-pleasant feeling or co-occurrence of low pleasure and high arousal consist of the Stroop interference effect. Therefore, dissociation of these two emotional dimensions is necessary to clarify emotional influence on the AMST.

The present experiment explored two questions; (1) whether unavoidable negative feelings contribute to memory impairment for the pairs of colour names-numbers on the AMST and (2) when an arousal level is neutral, whether the modification of the magnitude of the Stroop-related interference still occurs. To test these questions, this experiment aimed to study whether assigning good/bad values (i.e. non-arousal values) to numbers modifies recall memory accuracy of the word-number pairs with the AMST. To achieve this aim, the participants first learned the good/bad values of each number (e.g. "*1* is good and *5* is bad") and they studied the word-number pairs for the later cued-recall tests. To balance the emotional labels of the numbers, in Good-Bad group, the numbers *1* to *4* were good numbers and *5* to *8* were bad ones whilst these good/bad labels were reversed in Bad-Good group. The hypothesis was, if unavoidable negative feelings degrade memory performance and such feelings do not include the elevation of arousal, studying "bad" numbers in the incongruent condition would mitigate interference and reveal higher memory accuracy than learning "good" numbers in this condition.

Method

Participants

Sixty students of University of Kent at Canterbury were recruited as the participants (i.e. Good-Bad group and Bad-Good group consisted of 30 participants, respectively). Since the current word type (i.e. colour names) was different from Experiment 5-A and 5-B (i.e. emotional words), the prior participation experience for these two experiments was not considered whilst all participants had no prior experience of participating in Experiment 1-4 and 6-8. In Good-Bad group there were two male participants (*M* = 19.00 years old; *SD* = .000) and 28 female participants (*M* = 18.75 years old; *SD* = 0.75). Bad-Good group consisted of four male participants (*M* = 19.50 years old; *SD* = 1.00) and 26 female participants (*M* = 19.15 years old; *SD* = 1.16). This group of 60 participants consisted of 45 native English speakers (Good-Bad: *n* = 25; Bad-Good: *n* = 20), 12 non-native English speakers (Good-Bad: *n* = 4; Bad-Good: *n* = 8) and three "English plus other language(s)" speakers (Good-Bad: *n* = 1; Bad-Good: *n* = 2).

Design

The present experiment was a 3 (Colour Congruency; congruent, incongruent, neutral) x 3 (Study-Test Session; 1st, 2nd, 3rd) x 2 (Number Type; good, bad) x 2 (Finger Group; Good-Bad, Bad-Good) factorial design with three within-participant factors (i.e. Colour congruency, Study-test session, Number type) and one between-participant factor (i.e. Finger group). The assignments of colour congruency conditions and of finger groups were counterbalanced based on Williams Latin Square Design (Williams, 1949).

Materials

The materials were the same as Experiment 1 (see Appendix A) with the same ink colour congruency check questionnaire that was used in Experiment 5-B, 7 and 8.

Procedure

The majority of the procedure was identical to Experiment 1 except the number-categorisation tasks and the ink colour congruency check questionnaire as Experiment 5-B, 7 and 8 included. After the participants finished the 48-trial typing practice section, the screen asked them to remember the values of the numbers (i.e. good or bad), required response keys (i.e. “a” or “p”) for each number and finger positions for these keys (i.e. the left and right hands’ index fingers; see Figure 4.1.1). The picture of finger placement was shown permanently during the number-categorisation tasks. For the assignment of the good/bad values to each hand, it might be problematic if the good numbers were always associated with the left hand whilst the bad numbers were paired with the right hand due to the potential issue of dominant hands. To remove this potential confounding effect, the value-hand associations were counterbalanced by switching right and left (i.e. assigned to one of two groups: Good-Bad group and Bad-Good group).

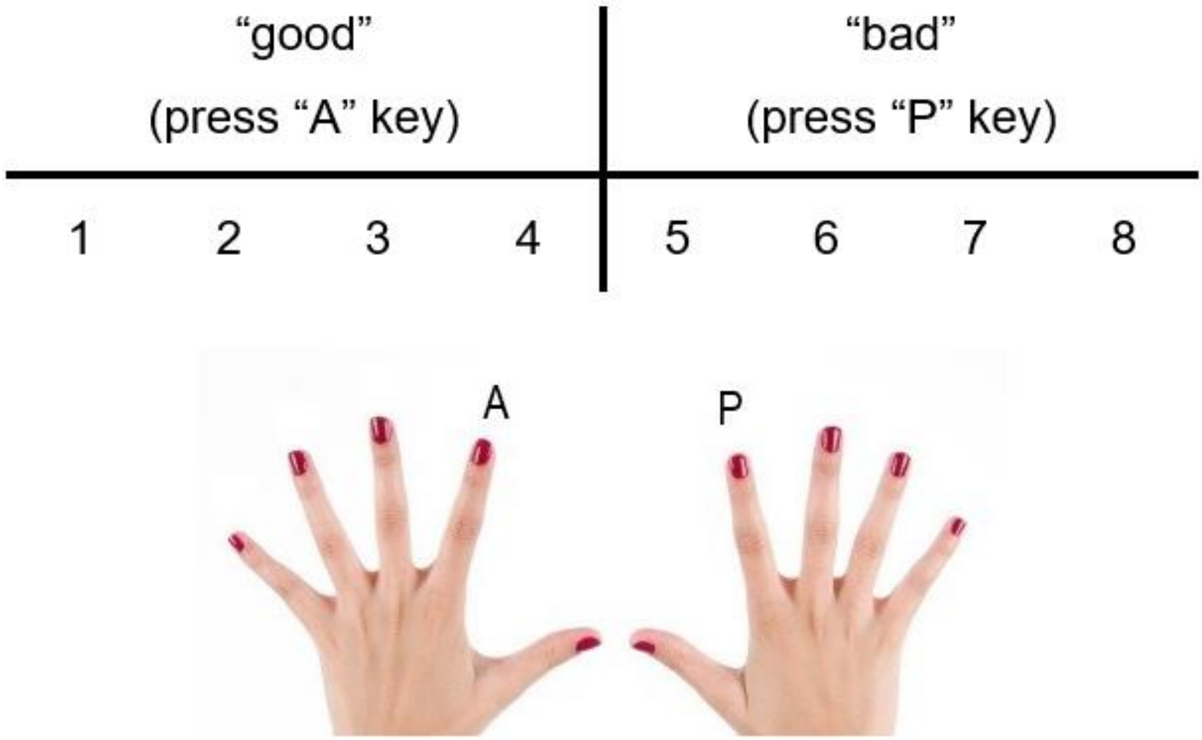


Figure 4.1.1. The picture of the screen the participants in Good-Bad group saw. The picture of two hands were obtained from the website (“typingMe.com,” n.d.) and the researcher added the alphabets “A” and “P” above both hands’ index fingers in order to show the finger placement. They were asked to remember these good/bad numbers, assigned response keys for each number and finger positions. In Bad-Good group the number *1* to *4* were the bad numbers and *5* to *8* were the good numbers.

When the participants felt confident about remembering these pieces of information, they were told to press the space key and the number-categorisation task started. During this task, the screen presented the number *1* to *8* (in black ink with the same font, size and position as the main to-be-studied numbers) one at a time in the random order and the participants were required to categorise whether it was a good or bad number by pressing the assigned key. There was no time limit for this task. The 500-ms feedback was provided to inform them whether their response was correct or incorrect. They repeated this 8-trial two times. This 16-trial number-categorisation task was inserted before each study-test session. Therefore, the present experiment consisted of 144 number-categorisation trials (16 trials x 3 sessions x 3 colour congruency conditions), 72 cued-recall test trials (8 word-number pairs x 3 sessions x 3 colour congruency conditions), 32 tone-comparison trials (16 trials x 2) and one ink colour congruency check questionnaire trial. This took approximately 30 minutes to complete.

Results

The following analyses focused only on performance on the cued-recall tests due to high mean correct proportions in all number-categorisation tasks [1st session, good (M = .97, SD = .04) and bad (M = .97, SD = .06); 2nd session, good (M = .97, SD = .04) and bad (M = .97, SD = .05); 3rd session, good (M = .97, SD = .04) and bad (M = .98, SD = .04)]. With the mean correct responses on the number-categorisation tasks, the 3 (Session) x 2 (Number Type) ANOVA with both factors as within-participants confirmed that the participants remembered the emotional labels of each number at equal level of accuracy in all three sessions [Session: $F(2, 118) = 0.41$, $p = .664$, partial-η^2 = .01; Number type: $F(1, 59) = 0.02$, $p = .889$, partial-η^2 = .000; Session x Number type: $F(2, 118) = 0.62$, $p = .542$, partial-η^2 = .01]. This could indicate that their memory accuracy for the type of numbers would have little impact on the following analyses. Note that four participants in Bad-Good group responded less than 200 ms; however, each of them made such a quick response only in one trial out of 144 trials. Therefore, it could be possible to confirm that all participants appropriately paid attention to all number-categorisation tasks.

Analysis of Accuracy

When analysing performance on the cued-recall tests, none of them was timed out and three participants in Good-Bad group made responses in less than 200 ms. However, since each of them made such a response only in one trial out of 72 trials, they were still considered as paying appropriate attention to the task and their responses were also included in the following analyses. The one-way ANOVA with Group as a between-participant factor revealed the identical memory accuracy between Good-Bad and Bad-Good groups [$F(1, 58) = 0.23$, $p = .635$, partial-$\eta^2 = .004$], which confirmed that memory accuracy was the same when the number *1* was a good number or a bad number. Thus, this group difference was not further analysed. Mean proportion correct responses and mean reaction times to good/bad numbers in the three sessions of the cued-recall tests are summarised in Table 4.1 and Figure 4.1.2.

Table 4.1.

Mean Proportion Correct Responses and Reaction Times (in ms) on the Cued-Recall Tests to the Good and Bad Numbers in Each Session of Each Colour Congruency Condition (Experiment 9)

Condition	Number Type	Mean Proportion Correct			Reaction Time		
		1st	2nd	3rd	1st	2nd	3rd
		M (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)
Congruent	Good	.49 (.28)	.66 (.25)	.76 (.29)	2,673 (1,100)	2,543 (873)	2,214 (1,002)
	Bad	.45 (.26)	.64 (.29)	.76 (.25)	3,179 (2,104)	2,627 (1,209)	2,345 (1,253)
Incongruent	Good	.30 (.26)	.47 (.30)	.54 (.34)	2,970 (1,519)	2,640 (1,155)	2,743 (1,360)
	Bad	.31 (.25)	.45 (.29)	.47 (.30)	2,892 (1,041)	3,086 (1,738)	2,959 (1,940)
Neutral	Good	.33 (.28)	.54 (.27)	.67 (.27)	2,933 (1,738)	2,674 (974)	2,451 (1,026)

	Bad	.42 (.25)	.54 (.30)	.69 (.27)	2,903 (1,391)	2,693 (1,265)	2,232 (1,077)

Note. The values represent performance when responding to the word-number pairs that included the good numbers or the bad numbers.

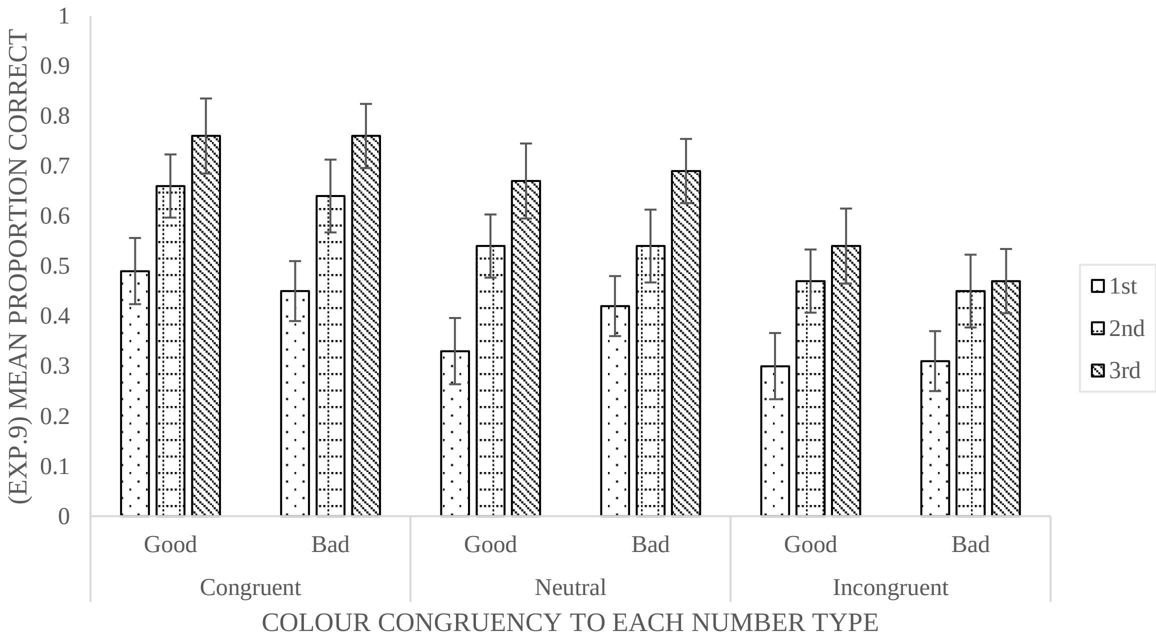


Figure 4.1.2. The graph shows the mean proportions of correct responses to the good and bad numbers in each colour congruency condition (Experiment 9). Error bars represent the 95% confidence intervals (Masson & Loftus, 2003).

The 3 (Colour Congruency) x 3 (Study-Test Session) x 2 (Number Type) factorial ANOVA was conducted with all factors as within-participant factors (see Table 4.1 and Figure 4.1.2).The analysis revealed the significant main effect of Colour congruency [$F(2, 118) = 21.32$, $p < .001$, partial-$\eta^2 = .27$] and of Study-Test session [$F(2, 118) = 110.82$, $p < .001$, partial-$\eta^2 = .65$] whilst other effects were not significant [Number type: $F(1, 59) = 0.07$, $p = .790$, partial-$\eta^2 = .001$; Colour congruency x Study-Test session: $F(4, 236) = 2.15$, $p = .076$, partial-$\eta^2 = .04$; Colour congruency x Number type: $F(2, 118) = 1.73$, $p = .182$, partial-$\eta^2 = .03$; Study-Test session x Number type: $F(2, 118) = 0.03$, $p = .513$, partial-$\eta^2 = .01$; Colour congruency x Study-test session x Number type: $F(4, 236) = 1.53$, $p = .195$, partial-$\eta^2 = .03$]. Paired sample *t*-tests (two-tailed) indicated the facilitation effect (Congruent: $M = .63$, $SD = .18$; Neutral: $M = .53$, $SD = .18$) [$t(59) = 2.92$, $p = .005$, $d = 0.38$] as well as the interference effect (Incongruent: $M = .42$, $SD = .20$) [$t(59) = 3.50$, $p = .001$, $d =$

0.45]. The difference between the congruent and incongruent conditions were also significant [t(59) = 6.84, p < .001, d = 0.88]. For the main effect of Study-Test session, the same post hoc tests indicated that the participants' scores on the cued-recall tests significantly increased as the sessions repeated [from the first session (M = .38, SD = .13) to the second (M = .55, SD = .15), t(59) = 9.08, p < .001, d = 1.17, and from the second to the third (M = .65, SD = .16), t(59) = 6.03, p < .001, d = 0.78].

Whilst the interaction effect of Colour congruency x Study-test session was not significant, as Experiment 1, the effects of colour-word congruency was examined. The 2 (Colour congruency; Congruent and Incongruent) x 3 (Study-Test Session) factorial ANOVA was conducted. This showed the same pattern of the results; the significant main effect of Colour congruency [F(1, 59) = 46.72, p < .001, partial-η^2 = .44] and of Study-test session [F(2, 118) = 72.62, p < .001, partial-η^2 = .55] but no significant interaction effect [F(1.818, 107.277) = 2.52, p = .090, partial-η^2 = .04]. These replicated the results in Experiment 1.

Analysis of Awareness

When considering the effect of the participants' awareness on memory accuracy with the good and bad numbers, the interaction of Colour congruency x Session x Number type was examined by the 3 (Colour Congruency) x 3 (Study-Test Session) x 2 (Number Type) x 2 (Ink Awareness) factorial ANOVA with Ink awareness as a between-participant factor whilst 33 participant correctly recognised all types of ink colour congruency. Since the analysis revealed the insignificant interaction effect of Colour congruency x Study-Test session x Number type x Ink awareness [F(4, 232) = 0.27, p = .900, partial-η^2 = .01], it would be possible to conclude that ink colour awareness had little impact on the reported memory accuracy.

Discussion

The present experiment tested whether unavoidable negative feelings impairs memory accuracy with the AMST and whether the changes in the robustness of the Stroop-related effects still occurs without the increment of arousal. This experiment was aimed to study whether good/bad values of numbers mitigates the interference effects from the incongruent ink colours when studying the word-number pairs on the AMST. Therefore, the hypothesis was that, in the incongruent condition, studying bad numbers would enhance cued-recall memory accuracy compared to studying good numbers in this condition. In the current experiment, after the participants studied the good/bad values of the numbers, they were asked to remember the pairs of colour names and numbers on the AMST and later these memories were tested on cued-recall tests. Their response accuracy on these memory tests was predominantly analysed and the analyses revealed that assigning the good/bad values to numbers had no impact on general memory accuracy with colour congruency; facilitation and interference were observed with the good and bad numbers. Therefore, the hypothesis was not supported and the outcomes contradicted Hatukai and Algom's (2017) study.

Such a contradiction might be due to a limited comparison of task performance in Hatukai and Algom's experiments. Specifically, the effects that Hatukai and Algom reported might occur only when comparing the good and bad responses and when colour congruency is relevant to the task. When they observed the mitigation of the Stroop interference effect, they compared performance only with the good and bad responses, not compared with the standard colour naming responses. For instance, their Figure 8 (i.e. the bar graph of the results in Experiment 7) illustrated that the mean reaction times when articulating "bad" to the incongruent Stroop stimuli was approximately 950 ms and saying "good" to these stimuli was close to 1,150 ms. Thus, they concluded that saying "bad" facilitates the response speed to this type of stimuli. Despite their conclusion, the reduction of the reaction time with a "bad" response seemed to occur only when comparing good and bad response. Although they considered the reaction time of 950 ms as supporting evidence for the reversal Stroop interference effect, this reaction time appeared to be larger than when the participants responded to the incongruent Stroop stimuli on the traditional Stroop task. Hatukai and Algom's Figure 3 (i.e. the bar graph of the results in Experiment 3) demonstrated that the mean reaction time in the incongruent condition was approximately 900 ms on the standard colour-naming task. This may raise a question of whether senses of "good" and "bad" are the major emotions that the Stroop objects induce from participants. Since there was no emotionally-neutral condition in Hatukai and Algom's research, it might be possible that saying "bad" to the incongruent Stroop stimuli made the participants' responses slower than the standard colour-naming responses to this type of stimuli. If this was the case, saying "bad" may be related to colour incongruency whilst this emotion may not be the major emotion that conflict from the incongruent Stroop stimuli induces. Because of this weak influence of good/bad feelings, colour congruency may require to be a centre of attention in order to produce effects on task performance. Since ink colours are to-be-ignored information on the AMST, feeling "good" and "bad" had no impact on the robustness of ink colour effects in the present experiment.

Additionally, the present experiment revealed the facilitation effect in all study-test sessions whilst this type of effects was observed only in the second and third sessions in Experiment 1. This different onset of facilitation might indicate how much practice was provided before the main task. Specifically, it may suggest that giving more practice before the main task induces facilitation earlier in the sessions. The participants in the present experiment underwent the number-categorisation trials before the main task. Although this was different from the main task (i.e. remembering the word-number pairs on the AMST), the additional exposure to the numbers may have helped the participants to feel familiar with the experiment faster or more robustly. Therefore, the occurrence of facilitation from the first session may signal that the timing of such an effect involves the amount of practice before the main task when such an effect on the AMST is measured by cued-recall tests. Compared to the stability of interference, this different timing of facilitation would suggest different underlying processes contribute to facilitation and interference; this can be something that future research can explore.

Overall, the present experiment demonstrated that assigning the “good” and “bad” values to the numbers failed to modify the robustness of ink colour effects when such effects are measured by cued-recall memory tests. However, since these emotional labels might not be influential enough to modify the robustness of the colour-word associations, the next experiment employed the emotional stimuli that activate emotions that conflict seems to induce: happy, sad and neutral face icons.

Experiment 10

The present experiment employed more distinctive emotional stimuli to examine the emotional effects of colour congruency on the AMST. The main assumptions in this experiment were that semantic conflict triggers emotional reactions (in Chapter 1) and dealing with task-irrelevant emotional information increases extraneous cognitive load (in Chapter 4 Introduction). Experiment 9 demonstrated that assigning the good and bad impressions to the numbers had no impact on the robustness of ink colour effects when such effects are measured by memory tests. However, due to the vague relationships between feelings of “good” and “bad” and the Stroop effect, the potential scenario emerged; the mitigation of the Stroop interference effect occurs only when using the good/bad labels and when colour congruency is task-relevant. To overcome this issue of emotional stimuli, the present experiment used new emotional stimuli that may be robust enough to modify the strength of conflict as van Steenbergen, Band and Hommel (2009) demonstrated using schematic emotional faces.

The study conducted by van Steenbergen et al. demonstrated the method to modify the magnitude of the semantic congruency effects with emotional stimuli. The researchers not only successfully modified the robustness but also analysed the direction of the effects by employing the emotionally-positive, -neutral and -negative stimuli with the notion that conflict induces aversiveness (Botvinick, 2007). Therefore, their research design seemed to illustrate the ideal methodology for the present experiment to follow. van Steenbergen et al. conducted research to examine whether conflict induces an aversive feeling. Their research idea was derived from the three notions: (1) Botvinick, Braver, Barch, Carter and Cohen’s (2001) finding of the conflict-adaptation effect (i.e. the function of cognitive control in order to adjust to conflict), (2) Botvinick’s (2007) report that conflict is experienced as an aversive event and (3) the literature (e.g. Dreisbach & Goschke, 2004) that demonstrated the facilitative effects from positive moods on cognitive control. By combining these, the researchers questioned whether the conflict signals aversiveness, this negative feeling might be a trigger for the conflict-adaptation effect and if so, such an adaptation process may be counteracted by being in a positive mood. Thus, their study aimed to test whether involving emotional events (i.e. monetary gain/loss) during the arrow flanker task would modify task performance. The hypothesis was that when the rewarding stimulus signalled monetary gain immediately after the incongruent stimulus, this would prevent the conflict adaptation in the next trial. The task was the typical arrow flanker task, which asked participants to judge the direction of the central arrow whilst ignoring the surrounding arrows. Specifically, when the central arrow and the surrounding arrows were facing in the same direction (e.g. all arrows pointed right), this was the congruent trial. When the central arrow was facing in a different direction from the surrounding arrows, this was the incongruent trial. The important manipulation of this study was that the three types of face icons were inserted between each trial; smiley faces indicated monetary gain, sad faces represented monetary loss and neutral faces illustrated no monetary changes. Note that the order of these face stimuli was randomised and these stimuli did not require a response. van Steenbergen et al. analysed the reaction times to the target arrows and observed the conflict-adaptation effect only with the sad and neutral faces; experiencing monetary gain reduced the occurrence of such an adaptation. Thus, they concluded that reward can eliminate conflict.

One question that the study conducted by van Steenbergen et al. (2009) may suggest would be the necessity of monetary gain/loss to modify the robustness of the Stroop effect. For example, winning money elevates gamblers' level of heart rate arousal (e.g. Wulfert, Roland, Hartley, Wang, & Franco, 2005). Thus, the use of both face icons and money would make it difficult to determine whether a low-pleasant feeling or the combination of low pleasure and high arousal consists of the negative emotion that conflict triggers. To dissociate these two dimensions, examination of the Stroop-related tasks with face icons would be appropriate as a starting point. Therefore, the present experiment used the face icons following van Steenbergen et al. (2009).

Experiment 10 explored the research question; if conflict information induces negative emotions and this generates memory impairment, whether such interference can be mitigated with the exposure to positive-valenced stimuli. Therefore, the aim was to examine whether presenting the three types of face icons (i.e. happy, sad and neutral) between the word-number pairs would modify task-irrelevant ink colour effects on the AMST when these effects are measured by the cued-recall tests. To achieve this aim, the participants were asked to remember the word-number pairs for the later cued-recall tests when the to-be-ignored happy, sad and neutral face icons were presented between each pair. The hypothesis was that the participants who saw the happy face icons in the incongruent condition would reveal higher memory test scores than those who saw the sad and neutral face icons.

Method

Participants

Thirty students of University of Kent at Canterbury were recruited as the participants. Since the current word type (i.e. colour names) was different from Experiment 5-A and 5-B (i.e. emotional words), the prior participation experience for these two experiments was not considered whilst all participants had no prior experience of participating in other experiments. There were three male participants (*M* = 20.00 years old; *SD* = 2.65) and 27 female participants (*M* = 18.96 years old; *SD* = 0.71) in Experiment 10. This group of 30 participants consisted of 22 native English speakers, seven non-native English speakers and one “English plus other language(s)” speaker.

Design

The present experiment was a 3 (Colour congruency; congruent, incongruent, neutral) x 3 (Session; 1st, 2nd, 3rd) x 3 (Face Icon Type; happy, sad, neutral) factorial design with all factors as within-participant factors. The order of Face icon type was counterbalanced in each colour congruency condition. For example, in the first colour congruency condition, some participants saw the happy faces in the first session whilst some of them saw those in the second session or the third session (see Appendix N).

Materials

The materials were identical to Experiment 9 (see Appendix A) except the points that the present experiment did not assign any affective valence to the learning materials; however, inserted face icons between each word-number pair (see Appendix N). There were three types of these icons: happy, sad and neutral. The face icons were obtained from the list of symbols in Microsoft. Since the category of Wingdings has three types of face icons (i.e. ☺, ☹ and 😐), these symbols were used. All face icons were presented in Wingdings font, 0.4 point of letter heights (i.e. bigger points indicate a bigger letter size), in black ink and at the centre of the screen in Psychopy (Peirce et al., 2019). These face icons were presented between each word-number pair for 500 ms and the inter-stimulus interval between the word-number pairs and the face icons was 250 ms.

Procedure

The majority of the procedure was identical to Experiment 9 except the task instruction. When the screen presented the task instruction after the practice section, the new instruction (i.e. “Please ignore face icons!”) was added. Since no responses were required to the face icons, the present experiment consisted of 72 cued-recall test trials (8 word-number pairs x 3 sessions x 3 colour congruency conditions), 32 tone-comparison trials (16 trials x 2) and one ink colour congruency check questionnaire trial. This took approximately 20 minutes to complete.

Results

The descriptive was summarised in Table 4.2.1 and Figure 4.2. None of the participants was timed out whilst one participant responded in less than 200 ms. However, since such a quick response was made one out of 72 cued-recall test trials, it could still indicate this participant's appropriate attention to the task. Therefore, all participants' responses on the cued-recall tests were included in the following analyses.

Analysis of Accuracy

For the effects of the face icon order, the one-way ANOVA with Order of face icons as a between-participant factor showed that memory accuracy was identical regardless of the face icon order [$F(5, 24) = 0.30$, $p = .910$, partial-$\eta^2 = .06$]; seeing the happy face icons in the first session or the last session did not change memory performance. Therefore, this factor was not further analysed.

Table 4.2.1

Mean Proportion Correct Responses and Reaction Times (in ms) on the Cued-Recall Tests with the Three Face Icon Types in Each Session of Each Colour Congruency Condition (Experiment 10)

Condition	Face Icon Type	Mean Proportion Correct			Reaction Time		
		1st	2nd	3rd	1st	2nd	3rd
		M (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)
Congruent	Happy	.55 (.24)	.66 (.32)	.80 (.25)	2,289 (497)	2,999 (1,135)	2,379 (658)
	Sad	.48 (.28)	.76 (.25)	.81 (.22)	3,121 (602)	2,128 (888)	1,841 (642)
	Neutral	.50 (.27)	.66 (.22)	.83 (.25)	2,916 (1,049)	2,749 (705)	2,266 (1,707)
Incongruent	Happy	.26 (.17)	.63 (.36)	.74 (.31)	2,181 (320)	2,737 (1,174)	2,545 (838)
	Sad	.43 (.35)	.56 (.23)	.59 (.37)	3,656 (1,724)	2,688 (541)	2,195 (885)
	Neutral	.34 (.26)	.58 (.35)	.74 (.24)	3,134 (604)	3,090 (2,930)	2,327 (1,415)
Neutral	Happy	.49 (.19)	.64 (.27)	.69 (.24)	2,153 (574)	2,325 (947)	3,046 (1,446)
	Sad	.45 (.31)	.53 (.34)	.60 (.26)	2,932 (873)	2,393 (855)	2,293 (926)
	Neutral	.40 (.28)	.55 (.20)	.68 (.28)	2,876 (760)	2,619 (908)	2,458 (1,401)

Note. Each cell represents the means with 10 participants who saw each face icon type in each session. For example, the upper left cell in the congruent condition indicates the mean correct proportion of the individuals who saw the happy faces in the first session in the congruent condition (i.e. those who were in the Happy-Sad-Neutral and Happy-Neutral-Sad groups). The next cell on the right represents the mean correct proportion of the participants who saw the happy faces in the second session in the congruent condition (i.e. those who were in the Neutral-Happy-Sad and Sad-Happy-Neutral groups).

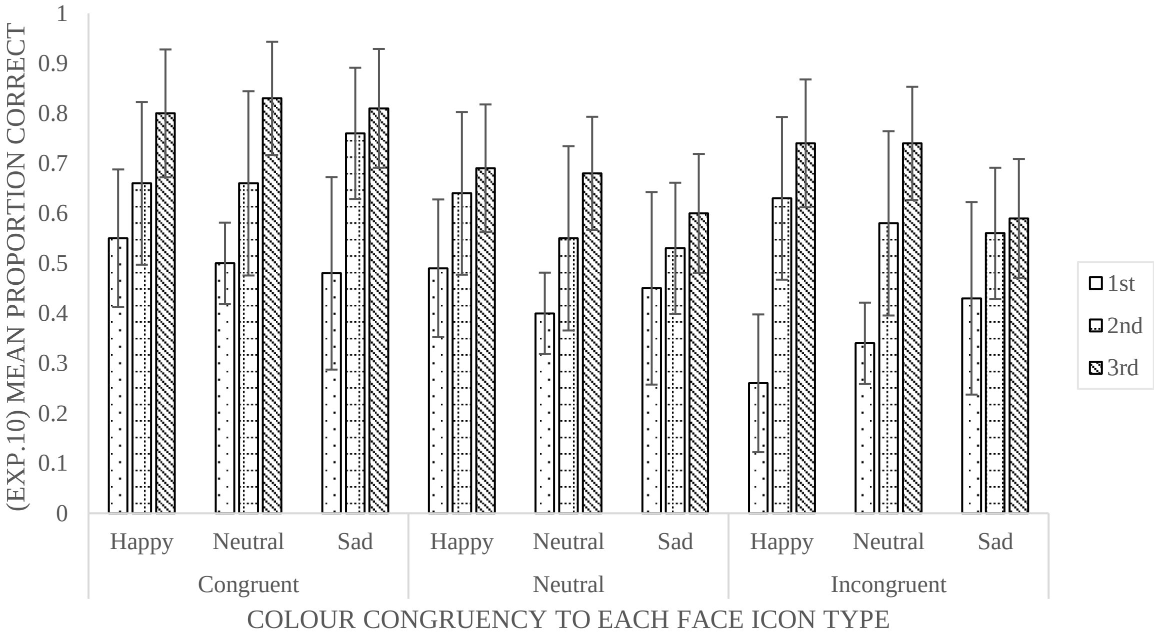


Figure 4.2. The graph shows the mean proportions of correct responses with the three face icon types in each session of the three colour congruency conditions (Experiment 10). Error bars represent the 95% confidence intervals (Masson & Loftus, 2003). Each bar represents the means with 10 participants who saw each face icon type in each session. For example, the bar on the furthest to the left in the congruent condition indicates the mean correct proportion of the individuals who saw the happy faces in the first session in the congruent condition (i.e. those who were in the Happy-Sad-Neutral and Happy-Neutral-Sad groups).

As some participants saw the happy face icons in the first session whilst others saw these in the second or third session, this analysis excluded the Study-test session factor since each face icon type could indicate the number of sessions that could reduce the complexity of the analyses. Therefore, the 3 (Colour Congruency) x 3 (Face Icon Type) factorial ANOVA was conducted with both as within-participant factors. The analysis revealed the significant main effect of Colour congruency [$F(2, 58) = 5.65$, $p = .006$, partial-$\eta^2 = .16$] whilst other effects did not reach at the significant level [Face icon type: $F(2, 58) = 0.18$, $p = .839$, partial-$\eta^2 = .06$; Colour congruency x Face icon type: $F(4, 116) = 0.77$, $p = .544$, partial-$\eta^2 = .03$]. For the main effect of Colour congruency, paired sample *t*-tests (two-tailed) indicated the significant facilitative effect [Congruent: $M = .67$, $SD = .22$; Neutral: $M = .56$, $SD = .20$; $t(29)$

= 3.05, p = .005, d = 0.56] and the significant difference between the congruent and incongruent condition [Incongruent: M = .54, SD = .26; $t(29)$ = 2.80, p = .009, d = 0.51] but no interference [$t(29)$ = 0.42, p = .679, d = 0.08]. When dropping the neutral condition, the 2 (Colour congruency) x 3 (Face Icon Type) factorial ANOVA demonstrated the same pattern of the results; the significant main effect of Colour congruency [$F(1, 29)$ = 7.85, p = .009, partial-η^2 = .21] but no significant main effect of Face icon type [$F(2, 58)$ = 0.001, p = .999, partial-η^2 = .00] and no significant interaction effect [$F(2, 58)$ = 0.31, p = .735, partial-η^2 = .01].

As the previously-reported experiments in the present thesis, the 3 (Colour congruency) x 3 (Study-Test Session) factorial ANOVA was performed with both as within-participant factors. The analysis indicated the same significant main effect of Colour congruency as well as the significant main effect of Study-test session [$F(1.638, 47.515)$ = 45.47, p < .001, partial-η^2 = .61]. As expected, paired sample t-tests (two-tailed) showed the significant increment from the 1st study-test session (M = .43, SD = .19) to the second session (M = .62, SD = .20) [$t(29)$ = -5.58, p < .001, d = 1.02] and from the second to the third (M = .72, SD = .22) [$t(29)$ = -4.50, p < .001, d = 0.82]. This pattern of the result was the same as the previously-reported experiments in the present thesis. Even when comparing only the congruent and incongruent conditions with the 2 (Colour congruency) x 3 (Study-Test Session) factorial ANOVA, this replicated the results; the significant main effect of Colour congruency [$F(1, 29)$ = 7.85, p = .009, partial-η^2 = .21] and of Study-test session [$F(1.579, 45.788)$ = 63.08, p < .001, partial-η^2 = .69] but no interaction effect [$F(2, 58)$ = 0.52, p = .595, partial-η^2 = .02].

Analysis of Awareness

When considering the effects of the participants' awareness of ink colour congruency on memory accuracy with the face icons, the 3 (Colour Congruency) x 3 (Face Icon Type) x 2

(Ink Awareness) factorial ANOVA with Ink awareness as a between-participant factor whilst 17 participants correctly recognised all types of ink colour congruency. This revealed that the interaction effect of Colour congruency x Face icon type x Ink awareness was not significant [$F(4, 112) = 2.44$, $p = .051$, partial-$\eta^2 = .08$]. This could indicate that whether the participants became aware of ink colour congruency or not had little impact on the reported memory performance with the face icon manipulation.

Comparison between Experiment 1 and 10

The reported analysis of accuracy demonstrated the disappearance of interference whilst this effect has been consistently observed with colour names in the present research. Two possible accounts might explain such a missing effect of interference; memory accuracy was poorer in the neutral condition or memory performance was enhanced in the incongruent condition compared to the previously-reported experiments. To test these possibilities, the memory test scores in the neutral and incongruent conditions were compared with Experiment 1 since this study also employed the colour names and the cued-recall tests with the three study-test sessions. This analysis aimed to examine what contributed to the missing interference effect; thus, the Study-test session factor was not considered. That is, the total number of correct responses in each colour congruency condition was divided by the total number of test trials and this total mean proportion correct responses were used in the following analysis. Two independent sample *t*-tests (two-tailed) were conducted and demonstrated that memory accuracy in the neutral condition was identical between Experiment 1 and 10 [$t(58) = -0.91$, $p = .368$, $d = 0.23$] whereas this was significantly different in the incongruent condition [$t(58) = -2.47$, $p = .017$, $d = 0.64$] (see Table 4.2.2). This could indicate that memory accuracy in the incongruent condition was significantly enhanced in Experiment 10 ($M = .54$, $SD = .26$) compared to Experiment 1 ($M = .39$, *SD*

= .20). Since the difference between the neutral and incongruent conditions became smaller in the present experiment, this may have contributed to the disappearance of interference.

Table 4.2.2

Mean Proportion Correct Responses on the Cued-Recall Tests in Each Colour Congruency Condition between Experiment 1 and Experiment 10

	Experiment 1	Experiment 10
Condition	Total	Total
	M (*SD*)	*M* (*SD*)
Congruent	.59 (.22)	.67 (.22)
Incongruent	.39 (.20)	.54 (.26)
Neutral	.51 (.19)	.56 (.20)

The next question to examine was whether this enhanced memory performance in the incongruent condition was due to seeing the happy face icons or because of having an extra time between each word-number pair. If the former was the case, this could support the hypothesis whilst the latter would be consistent with the duration of study account suggested from Chapter 3. To test these assumptions, the total mean correct proportion in Experiment 1 was compared with when seeing the happy, sad and neutral face icons in Experiment 10. Three independent sample *t*-tests (two-tailed) showed significant differences between Experiment 1 and seeing the happy face icons in Experiment 10 (M = .54, SD = .35) [$t(46.738) = -2.07$, $p = .044$, $d = 0.53$] and the neutral face icons (M = .55, SD = .33) [$t(48.727) = -2.29$, $p = .026$, $d = 0.59$] whilst seeing the sad face icons (M = .53, SD = .32) also demonstrated the close-to-significant memory enhancement in Experiment 10 [$t(49.205) = -1.96$, $p = .056$, $d = 0.51$]. To examine whether the robustness of these memory performance changes differed depending on the type of face icons, the researcher subtracted the total mean proportion correct with the incongruent ink colours in Experiment 1 from the total mean proportion correct in the incongruent condition with the happy face icons in Experiment 11. The same subtraction was performed with the sad and neutral face icons. With these values, the one-way ANOVA with Face icon type was performed and revealed no main effect of Face icon type [$F(2, 58) = 0.08$, $p = .925$, partial-$\eta^2 = .003$]. This could indicate that the robustness of memory performance changes between Experiment 1 and 11 with the incongruent ink colours were identical when seeing the happy, sad or neutral face icons. Thus, seeing these three types of face icons generally enhanced memory accuracy in the incongruent condition compared to Experiment 1, which did not support the hypothesis of the present experiment.

Discussion

Experiment 10 explored, if the negative emotions that are triggered by colour-word incongruency on the AMST produce memory impairment, whether such a detrimental influence can be mitigated by presenting the positive-valenced stimuli. The aim of this experiment was to examine whether the face icons that appeared between each word-number pair would modify performance on the cued-recall memory tests in the incongruent condition. Therefore, it was hypothesised that seeing the happy face icons in the incongruent condition would mitigate interference on memory performance compared to the sad and neutral face icons. In this experiment the participants were asked to remember the pairs of colour names and numbers (e.g. *blue-4*) whilst ignoring the face icons (e.g. ☺) appeared between each pair. As expected, the facilitative effect was observed whereas interference disappeared in the present experiment. However, this disappearance of interference occurred with all three types of face icons, which did not provide support for the hypothesis.

Although the reported results were different from van Steenbergen et al.'s (2009) study, it could be explained by the weak valence values of the emotional stimuli. This was suggested due to a lack of the exact valence values of the face icons. Whilst these face icons represented each emotion (i.e. positive, negative and neutral) distinctively, it was not known how strongly these face icons induce each emotion from the participants. Therefore, it would be possible to argue that since the valence values of these face icons were not high enough to produce effects, the present experiment failed to observe the mitigation of interference only with the happy face icons. This issue would be solved by using the emotional stimuli that the valence values of these were examined previously.

To recap, Experiment 10 revealed that memory accuracy in the incongruent condition was enhanced regardless of the type of face icons on the AMST. However, since this might be due to the methodological issue in the present experiment (i.e. the valence value of the face icons), the next experiment considered these. Thus, Experiment 11 selected the auditory emotional stimuli (Bradley & Lang, 2007) that show the previously-measured valence values.

Experiment 11

The present experiment explored whether using the emotional sounds could produce any modification effects on the robustness of ink colour effects with the AMST when such effects are measured by memory tests. Whereas van Steenbergen et al. (2009) reported that seeing the smiley face icons, which indicated monetary gain (i.e. the positive stimuli), can mitigate the level of the induced negative feelings from conflict, Experiment 10 in this thesis revealed that seeing the three types of face icons (i.e. the happy, sad and neutral face icons) identically increased memory performance with the incongruent ink colours (i.e. a source of conflict). However, there was a methodological limitation in Experiment 10: the degree of affective valence of the face icons was missing. To overcome this, the present experiment selected the emotional sounds based on normative ratings by Bradley and Lang’s (2007), which include the pleasure and arousal values of the auditory stimuli (e.g. laughing). The hypothesis was identical to Experiment 10; presenting the positive sounds during study would mitigate the interference effects from incongruent ink colours on the cued-recall test.

The reason for changing the modality of the emotional stimuli was to prevent the overload of working memory whilst presenting additional to-be-ignored information. The separate working memory systems manage visual information and auditory one (A. Baddeley, 2012). Since the focus of the present experiment was memory performance of word-number pairs with colour congruency whilst ignoring the emotional stimuli, it was important to prevent the circumstance that processing these emotional stimuli induces the overload of working memory. That is, if the emotional pictures were presented during studying, inhibiting the processes of these stimuli might affect remembering the subsequent word-number pairs since these were both visual stimuli. This issue was of concern in the present experiment because the pictures are assumed to deliver more complex visual information than the face icons used in the previous experiment. Additionally, the use of the pictures would also have an issue of colouring. It would be questionable whether presenting a picture of a fun birthday party in black-and-white can induce the same degree of pleasure when it is presented with colours. However, presenting it with colours might interact with the ink colours of the numbers, which needed to be avoided. Therefore, the present experiment presented the emotional stimuli in an auditory format.

Method

Participants

Thirty students of University of Kent at Canterbury were recruited as the participants. Since the current word type (i.e. colour names) was different from Experiment 5-A and 5-B (i.e. emotional words), the prior participation experience for these two experiments was not considered whilst all participants had no prior experience of participating in other experiments. There were six male participants (*M* = 19.33 years old; *SD* = 1.21) and 24 female participants (*M* = 19.04 years old; *SD* = 0.96) in Experiment 11. This group of 30 participants consisted of 24 native English speakers, five non-native English speakers and one “English plus other language(s)” speaker.

Design

The design was the same as Experiment 10 except the factor, Face icon type, was replaced with Sound type and the face icons were replaced with the emotional sounds (i.e. positive, negative and neutral). The order of these sounds in each session was counterbalanced based on Experiment 10 (see Appendix N).

Materials

The materials were identical to Experiment 10 (see Appendix A) except that the face icons were replaced with the emotional sounds that were obtained from Bradley and Lang’s (2007) the International Affective Digitized Sounds 2nd Edition. They employed the 9-point pleasure and arousal scales (i.e. *9* indicated high on these dimensions) and reported the sound list with these values. Similar to the previous experiments, the ones that have the close-to-middle arousal scores, the close-to-extreme pleasure scores and the close-to-middle pleasure scores were selected. The selected emotional sounds, mean pleasure scores and arousal scores with each type of the emotional sounds were shown in Table 4.3.1.

Table 4.3.1

Descriptions and Sound Numbers used in Bradley and Lang's (2007) report, Mean Pleasure and Arousal Scores of Each Emotional Sound (Experiment 11)

Sound Type	Description	Sound Number	Pleasure	Arousal
			M (*SD*)	*M* (*SD*)
Positive	Applause1	351	7.40 (0.29)	5.48 (1.14)
	Baby	110		
	Bongos	817		
	BoyLaugh	220		
	FunkMusic	820		
	Harp	809		
	Laughing	226		
	Robin	151		
Negative	BattleTaps	611	2.82 (0.38)	6.08 (0.61)
	Bees	115		
	BusySignal	703		
	Buzzing	116		
	Explosion	626		
	FemaleCough	242		
	Fight3	283		
	MaleCough	241		
Neutral	Dog	107	5.07 (0.32)	5.37 (0.59)
	Panting	104		
	Phone1	704		
	Rooster	120		
	Train	425		
	TypeWriter	322		
	Walking	722		
	Writing	358		

To confirm that the mean pleasure scores between each sound type were statistically different, the one-way ANOVA with Sound type as a between-participant factor was conducted and it confirmed that there was a significant difference between the three types of the emotional sounds, $F(2, 21) = 375.89$, $p < .001$, partial-$\eta^2 = .97$. Planned contrasts indicated that the three sound types were significantly different from each other on the pleasure dimension (all $p < .001$). The same one-way ANOVA confirmed no difference in the level of arousal, $F(2, 21) = 1.74$, $p = .200$, partial-$\eta^2 = .14$. The selected emotional sounds were presented between each word-number pair for 6,000 ms and the inter-stimulus interval between the word-number pairs and the emotional sounds was 250 ms.

Procedure

The procedure was the same as Experiment 10; 72 cued-recall test trials (8 word-number pairs x 3 sessions x 3 colour congruency conditions), 32 tone-comparison trials (16 trials x 2) and one ink colour congruency check questionnaire trial. This took approximately 25 minutes to complete.

Results

The descriptive are summarised in Table 4.3.2 and Figure 4.3. None of the participants was timed out and all responses took more than 200 ms to be made; therefore, all participants' responses on the cued-recall tests were included in the following analyses.

Analysis of Accuracy

To examine whether the order of the emotional sounds affected memory performance, the one-way ANOVA with Order of sounds as a between-participant factor was performed. This confirmed that memory accuracy was identical across all patterns of orders, [$F(5, 24) = 0.46$, $p = .805$, partial-$\eta^2 = .09$]; thus, this factor was not further analysed.

Table 4.3.2

Mean Proportion Correct Responses and Reaction Times (in ms) on the Cued-Recall Tests with the Three Sound Types in Each Session of Each Colour Congruency Condition (Experiment 11)

Condition	Sound Type	Mean Proportion Correct			Reaction Time		
		1st	2nd	3rd	1st	2nd	3rd
		M (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)
Congruent	Positive	.73 (.26)	.69 (.31)	.90 (.18)	3,331 (893)	3,464 (2,008)	1,952 (582)
	Negative	.58 (.33)	.83 (.24)	.93 (.16)	2,969 (1,084)	2,327 (802)	1,958 (618)
	Neutral	.52 (.18)	.75 (.26)	.90 (.13)	3,735 (1,247)	2,162 (616)	2,305 (1,030)
Incongruent	Positive	.41 (.24)	.61 (.34)	.91 (.10)	3,913 (1,200)	3,825 (2,417)	2,167 (805)
	Negative	.65 (.27)	.65 (.36)	.74 (.29)	3,424 (2,463)	2,746 (1,084)	2,900 (1,291)
	Neutral	.46 (.25)	.78 (.19)	.78 (.23)	3,631 (1,365)	2,720 (1,770)	2,489 (1,530)
Neutral	Positive	.56 (.28)	.75 (.19)	.89 (.19)	2,922 (893)	2,705 (787)	1,970 (585)
	Negative	.58 (.29)	.76 (.22)	.80 (.17)	3,389 (1,477)	2,659 (768)	2,354 (873)
	Neutral	.39 (.17)	.66 (.25)	.95 (.09)	3,700 (2,073)	2,124 (510)	2,144 (941)

Note. Each cell represents the means with 10 participants who heard each sound type in each session. For example, the upper left cell in the congruent condition indicates the mean correct proportion of the individuals who heard the positive sounds in the first session in the congruent condition (i.e. those who were in the Positive-Negative-Neutral and Positive-Neutral-Negative groups). The next cell on the right represents the mean correct proportion of the participants who heard the positive sounds in the second session in the congruent condition (i.e. those who were in the Neutral-Positive-Negative and Negative-Positive-Neutral groups).

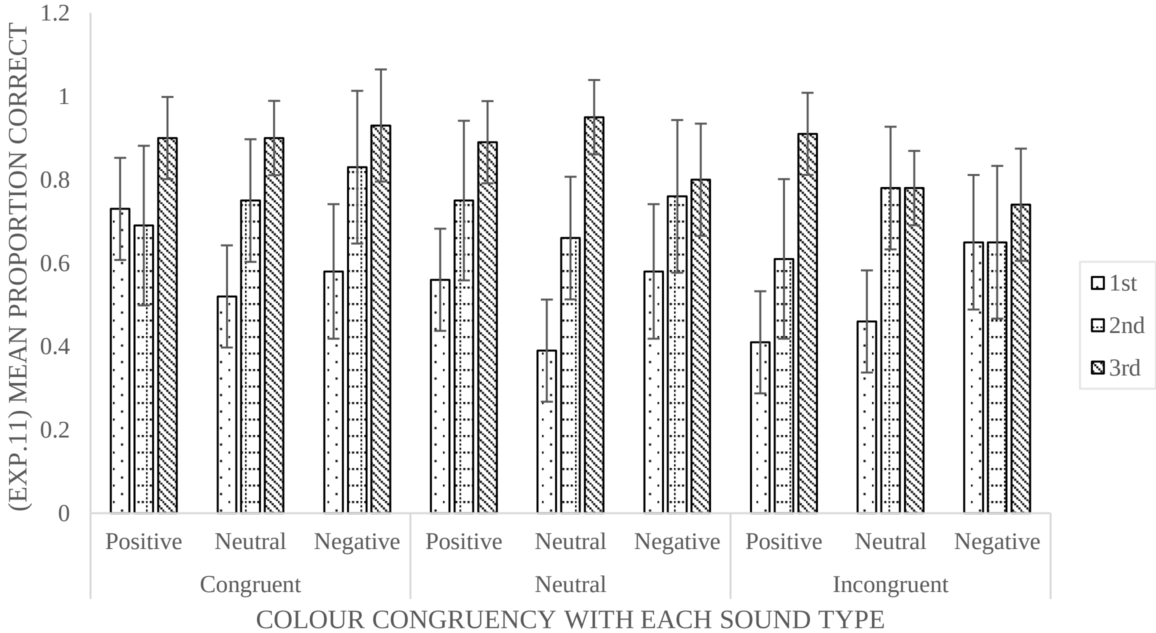


Figure 4.3. The graph shows the mean proportions of correct responses with the three sound types in each session of the three colour congruency conditions (Experiment 11). Error bars represent the 95% confidence intervals (Masson & Loftus, 2003). Each bar represents the means with 10 participants who heard each sound type in each session. For example, the bar on the furthest to the left in the congruent condition indicates the

mean correct proportion of the individuals who heard the positive sounds in the first session in the congruent condition (i.e. those who were in the Positive-Negative-Neutral and Positive-Neutral-Negative groups).

The 3 (Colour congruency) x 3 (Sound Type) factorial ANOVA with both factors as within-participant factors was conducted. Differing from the previous experiments, the analysis revealed that none of the effects reached at the significant level [Colour congruency: $F(2, 58) = 2.33$, $p = .106$, partial-$\eta^2 = .07$; Sound type: $F(2, 58) = 0.28$, $p = .759$, partial-$\eta^2 = .01$; Colour congruency x Sound type: $F(4, 116) = 0.58$, $p = .677$, partial-$\eta^2 = .02$]. Consequently, the analysis illustrated presenting the emotional sounds during study mitigated the robustness of task-irrelevant ink colour effects on the AMST; however, regardless of the type of the emotional sounds. When dropping the neutral condition, the 2 (Colour congruency) x 3 (Sound Type) factorial ANOVA replicated the results; no significant main effect of Colour congruency [$F(1, 29) = 3.26$, $p = .081$, partial-$\eta^2 = .10$], of Sound type [$F(2, 58) = 0.17$, $p = .846$, partial-$\eta^2 = .01$] and no interaction effect [$F(2, 58) = 0.63$, $p = .534$, partial-$\eta^2 = .02$].

The 3 (Colour Congruency) x 3 (Study-Test Session) factorial ANOVA was performed with both as within-participant factors. The analysis indicated the significant main effect of Session [$F(1.575, 45.680) = 72.27$, $p < .001$, partial-$\eta^2 = .71$]. The interaction effect of Colour congruency x Study-test session was not significant [$F(4, 116) = 0.52$, $p = .725$, partial-$\eta^2 = .02$]. As expected, paired sample *t*-tests (two-tailed) showed the significant increment of the mean proportion correct from the 1st study-test session ($M = .54$, $SD = .20$) to the second session ($M = .72$, $SD = .16$) [$t(29) = -6.42$, $p < .001$, $d = 1.17$] and from the second to the third ($M = .87$, $SD = .12$) [$t(29) = -7.46$, $p < .001$, $d = 1.36$]. This pattern of the result was the same as the previously-reported experiments in the present thesis. When comparing only the congruent and incongruent conditions, the 2 (Colour congruency) x 3 (Study-Test Session) factorial ANOVA showed the same pattern of the result; no significant main effect of Colour congruency [$F(1, 29) = 3.26$, $p = .081$, partial-$\eta^2 = .10$] and no significant interaction effect [$F(2, 58) = 0.10$, $p = .903$, partial-$\eta^2 = .004$] but the significant main effect of Study-test session [$F(2, 58) = 42.03$, $p < .001$, partial-$\eta^2 = .59$].

Whilst the interaction effect of Colour congruency x Study-test session was not significant, the potential ceiling effect might have contributed to it. For example, in Table 4.3.2, the mean proportion correct with the positive emotional sounds in the congruent condition already reached at .73 in the 1st session. Thus, it would be possible to assume that since the memory test scores were already close to the highest scores from the first session, these were stable across the three sessions. To make the values in each session of each colour congruency condition comparable, memory performance in the congruent and incongruent conditions were normalised with the baseline condition (i.e. the neutral condition). For example, the mean proportion correct in the 1st session of the congruent condition was divided by in the 1st session of the neutral condition. With these comparable values, the 2 (Colour congruency; Congruent and Incongruent) x 3 (Study-Test Session) factorial ANOVA was conducted. However, again the analysis revealed no significant interaction effect, $F(2, 58) = 0.52$, $p = .600$, partial-$\eta^2 = .02$. Since the pattern of the results was the same between the original values and the normalised values, this could signal that the ceiling effect had little impact on the reported effects.

Analysis of Awareness

When considering the effect of ink colour congruency awareness on memory accuracy, the 3 (Colour Congruency) x 3 (Sound Type) x 2 (Ink Awareness) factorial ANOVA with Ink awareness as a between-participant factor was performed whilst 18 participants correctly recognised all types of ink colour congruency. This indicated that the interaction effect of Colour congruency x Sound type x Ink awareness was significant [$F(4, 112) = 2.85$, $p = .027$, partial-$\eta^2 = .09$]. Therefore, the two 3 (Colour Congruency) x 3 (Sound Type) factorial ANOVAs were performed with those who correctly recognised colour congruency (*n* = 18) and with those who did not (*n* = 12). Whilst all effects were insignificant with those who did not (all $p > .05$), the main effect of Colour congruency reached at near the significant level with those who correctly recognised [F(1.505, 25.582) = 3.51, $p = .057$, partial-η^2 = .17]. With these participants, other effects were not significant (all $p > .05$). However, paired sample *t*-tests (two-tailed) indicated that the significant memory enhancement occurred only between the congruent (*M* = .79, *SD* = .19) and incongruent conditions (*M* = .63, *SD* = .26) [t(17) = 2.11, $p = .050$, $d = 0.50$]; no facilitation and interference. Therefore, it would be possible to conclude that ink colour awareness had little impact on the reported memory accuracy with the emotional sound manipulation.

Comparison between Experiment 1 and 11

The analysis of accuracy demonstrated that memory accuracy in all colour congruency conditions was identical in the present experiment whilst the other experiments in the present thesis reported the facilitative and/or interference effects. For a missing of facilitation, there would be two possible cases; memory test scores in the congruent condition decreased or these in the neutral condition increased. Similarly, the disappearance of interference might have occurred due to decreasing the scores in the neutral condition or increasing these in the incongruent condition. To clarify how listening to the emotional sounds modified the robustness of each effect, three independent sample *t*-tests (two-tailed) were conducted with the total mean proportion correct in each colour congruency condition between Experiment 1 and 11 (see Table 4.3.2). These analyses illustrated that memory accuracy in all colour congruency conditions was significantly enhanced in Experiment 11 compared to Experiment 1, [Congruent: t(58) = -3.11, $p = .003$, $d = 0.80$; Incongruent: t(58) = -5.03, $p < .001$, $d = 1.30$; Neutral: t(58) = -4.22, $p < .001$, $d = 1.09$].

Table 4.3.2

Mean Proportion Correct Responses on the Cued-Recall Tests in Each Colour Congruency Condition between Experiment 1 and Experiment 11

Condition	Experiment 1	Experiment 11
	Total	Total
	M (*SD*)	*M* (*SD*)
Congruent	.59 (.22)	.76 (.20)
Incongruent	.39 (.20)	.67 (.22)
Neutral	.51 (.19)	.70 (.16)

Such memory enhancement may raise a question; whether listening to the emotional sounds equally improved memory accuracy in all colour congruency conditions between the two experiments or the robustness of such memory enhancement differed depending on the type of emotional sounds in each colour congruency condition. To examine this question, the researcher subtracted the total mean proportion correct with the congruent condition in Experiment 1 from the total mean proportion correct in the congruent condition with the positive sounds in Experiment 11. The same subtraction was performed with the negative and neutral sounds. This calculation of difference between Experiment 1 and each sound type in Experiment 11 was also conducted to the incongruent and neutral conditions. With these values, the three one-way ANOVAs with Sound type were performed. These analyses indicated that the main effect of Sound type was not significant in all three colour congruency conditions [Congruent: $F(2, 58) = 0.46$, $p = .636$, partial-$\eta^2 = .02$; Incongruent: $F(2, 58) = 0.16$, $p = .856$, partial-$\eta^2 = .01$; Neutral: $F(2, 58) = 0.53$, $p = .594$, partial-$\eta^2 = .02$]. These could indicate that memory performance changes in each colour congruency condition between Experiment 1 and 11 occurred equally when listening to the positive, negative or neutral sounds, which would reject the hypothesis but provide support for the duration of study account suggested from Chapter 3.

Another question might be whether the degree of this memory performance change between Experiment 1 and 11 differed depending on colour congruency. To test this, the difference of the total mean proportion correct between Experiment 1 and 11 in the three colour congruency conditions was analysed with the one-way ANOVA, which was conducted with Colour congruency as a within-participant factor. Although this analysis indicated no significant main effect of Colour congruency [$F(2, 58) = 2.36$, $p = .104$, partial-$\eta^2 = .08$], Table 4.3.2 demonstrated that memory accuracy in the incongruent condition was improved the greatest between the two experiments. When comparing this difference in the incongruent condition with the difference in the other two conditions via paired sample *t*-tests (two-tailed), it showed that the degree of memory performance difference in the incongruent condition across Experiment 1 and 11 was significantly greater than in the congruent condition between these two experiments, $t(29) = -2.04$, $p = .050$, $d = 0.37$. Thus, the analyses seemed to indicate that listening to the emotional sounds improved memory accuracy in all colour congruency conditions whilst the robustness of such memory enhancement seemed to differ between the congruent and incongruent conditions.

Discussion

The present experiment was conducted to overcome the methodological flaw of Experiment 10 in order to test the hypothesis; hearing the emotionally positive sounds (e.g. laughing) would mitigate the detrimental impacts of incongruent ink colours on accuracy of the cued-recall memory. To examine this, the participants were asked to remember the pairs of colour names and numbers on the AMST for the later cued-recall tests whilst ignoring the emotional sounds that were inserted between each word-number pair. The pivotal finding of the current experiment was that when such emotional sounds were presented during learning, the participants' memory accuracy became identical across all three colour congruency conditions regardless of the type of the emotional sounds. Therefore, the hypothesis was not supported; memory accuracy in all colour congruency conditions was enhanced by listening to the emotional sounds whilst such memory enhancement was more robust with the incongruent ink colours than with the congruent ink colours

The general memory improvement with the emotional sounds might have occurred due to the longer inter-stimulus interval than Experiment 1. Note that, in Experiment 11, the inter-stimulus interval referred to the total time between when the presented word-number pairs disappeared and the next one appeared on screen. Elaboration of encoded information tends to produce the stronger memory traces whilst this deeper analysis requires additional time (Craik & Lockhart, 1972). This would suggest that since there was 6,000-ms inter-stimulus intervals in the present experiment whereas there were no such intervals in Experiment 1, it would be possible to presume that having a longer time to process the encoded information of the learning materials enhanced memory accuracy in Experiment 11.

The disappearance of the interference effect from the incongruent ink colours could be also explained by the longer inter-stimulus interval. Specifically, due to having a longer time to overcome conflict by suppressing task-irrelevant information, the participants' memory accuracy in the incongruent condition was enhanced. Such a relationship between inhibition and duration on cognitive tasks was reported by Zamorano et al. (2014), who studied the temporal effects on behaviour between physiological measurement (e.g. fMRI or EEG). The task was the Go-Nogo task, which asked the participants to press the button when the Go stimuli appeared whilst no response was required for the Nogo stimuli. The researchers found the stronger relationship between the Nogo responses and the N2 component (i.e. an event-related potential component or a brain response that reflects behavioural inhibition) in the reduced inter-stimulus interval condition compared to the longer condition. In addition to this physiological evidence, they also reported behavioural analyses that response accuracy to the Nogo stimuli was significantly impaired in the reduced inter-stimulus interval condition in comparison with the longer condition. By referring Jodo and Kayama's (1992) study, Zamorano et al. (2014) interpreted such findings as, when the inter-stimulus interval was shorter, inhibition of responses became more demanding compared to the longer inter-stimulus interval condition. Zamorano et al. suggested that extended inter-stimulus intervals allow participants to perform cognitive control at a more sophisticated level and assist decision-making processes to occur naturally that enhances response accuracy. Therefore, the study conducted by Zamorano et al. provided both behavioural and physiological evidence that having longer spare time between the to-be-processed items helps participants' inhibition processes, which facilitates accuracy of task performance. Zamorano et al. mainly discussed response inhibition; however, as they connected the function of the longer inter-stimulus interval with cognitive controls, the present research suggests that their findings can be extended to a more general behavioural inhibition. In this vein, it would be possible to explain the result of the current experiment based on the finding reported by Zamorano et al. (2014). That is, the 6,000-ms inter-stimulus interval in the present experiment might have provided additional time to inhibit the processes of the task-irrelevant information, which may have mitigated the detrimental effects from the incongruent ink colours on the AMST. When considering the processes of additional information (i.e. the emotional sounds) during learning, one might question whether such processes have disturbed the participants to remember the word-number pairs. However, the literature reported that unless the stimuli presented during the inter-stimulus intervals required attention, such materials had no impact on memory accuracy; rather than what learners see during the inter-stimulus intervals, the duration is more influential on memory performance (Intraub, 1980).

This relationship between inter-stimulus intervals and suppression behaviour may also describe the lack of facilitation in the present experiment. When studying the word-number pairs, only the neutral condition and incongruent condition appeared to have involved the suppression of information. Specifically, in the neutral condition the participants needed to suppress the activated colour images by reading the colour names whilst in the incongruent condition they were required to suppress such information and the activated colour names by seeing the incongruent ink colours. Since the longer inter-stimulus interval seems to produce benefits on elaboration of information and suppression behaviour, these two conditions might have received more advantages from the longer inter-stimulus intervals than the congruent condition. Consequently, the difference in memory accuracy between the congruent and neutral conditions became smaller in Experiment 11 compared to Experiment 1. In addition to this, since the incongruent condition may have required more suppression processes than the neutral condition, memory accuracy with the incongruent ink colours might have increased the greatest. In this way, memory performance in all colour congruency conditions became identical in the present experiment; the duration of inter-stimulus intervals might have a larger impact on modifying the robustness of interference than of facilitation when ink colours on the AMST are measured by memory tests. This would suggest to consider the interaction effect of duration of processing learning materials and semantics of such materials when measuring performance on the AMST.

To sum up, Experiment 11 employed the emotional sounds with the AMST and found that the participants' cued-recall memory accuracy in all colour congruency conditions became identical when such stimuli were presented between the learning materials. Additionally, the facilitative and interference effects from the congruent and incongruent ink colours disappeared that may have occurred due to the longer inter-stimulus interval. Thus, the current experiment demonstrated evidence that even if additional task-irrelevant stimuli are inserted, exposure to these can generate the advantageous effects on memory performance by providing the additional time.

Chapter 4: Chapter Summary

Based on the results of Chapter 2, Chapter 4 explored why memory impairment occurred in the incongruent condition from the emotional perspective. Specifically, the present chapter examined whether negative emotions induced by ink colour incongruency on the AMST produce the detrimental effects on memory accuracy. Thus, Chapter 4 was aimed to clarify whether manipulation of the affective valence of the learning materials and/or of the participants' mood during the task would modify the magnitude of interference with the cued-recall tests. The overall hypothesis was that, in the incongruent condition, studying the negative-valenced materials and exposure to the emotionally negative distractors would mitigate interference due to increased mood congruency between learners and materials. To test this hypothesis, the participants were asked to remember the pairs of colour names and the "good" and "bad" numbers on the AMST in Experiment 9. In Experiment 10 and 11, the to-be-ignored emotional stimuli were presented between each word-number pair. Later, these memories were tested on the cued-recall tests. Through these three experiments, it was found that none of these emotional manipulations modified the robustness of task-irrelevant ink colour effects on the AMST. Regardless of the manipulations on the emotional dimension, the highest memory test scores in the congruent condition and poorest in the incongruent condition were observed in Experiment 9 and only the former result occurred in Experiment 10. However, this disappearance of interference occurred regardless of the type of the emotional stimuli and these facilitation and interference disappeared in Experiment 11. Therefore, the hypothesis was not supported when assigning the good/bad valences to the learning materials and presenting the emotional stimuli during study.

The different results in Experiment 9, 10 and 11 from the literature (Hatukai & Algom, 2017; van Steenbergen et al., 2009) might be explained by the temporal factor in each trial. Hatukai and Algom (2017) and van Steenbergen et al. (2009) reported that manipulation on the emotional dimension modified the response speed on the cognitive tasks. However, when no pressure was placed on the reaction time dimension as in the present research, such manipulations seemed to have no notable impact on response accuracy. From these findings, the present research would suggest a tentative explanation; task-irrelevant negative emotions induced by the conflicting information can be suppressed with the cost of the response speed. The basic assumption of this explanation was that conflicting stimuli activate negative emotions as Damen, Strick, Taris and Aarts (2018) reported. The point here would be why these emotions affected only on the speed dimension. When considering this point, the important difference between the literature (Hatukai & Algom, 2017; van Steenbergen et al., 2009) and the current research is whether the participants were required to respond immediately or not. In the literature making the immediate responses were highlighted. Whilst the present research also asked the participants to make the responses as fast as possible, accurate responses were also required and the participants were informed that 2 minutes were given to answer only 16 trials, which allowed them to respond carefully in order to make the accurate responses. Therefore, by informing the participants that they would have enough time to make responses, it was assumed that the participants' feeling of pressure on reaction times was different between the literature and the current experiments. This difference in the research design of time restriction may have contributed to the contradicting results between previous research and the studies presented here. That is, when the participants felt negative due to the conflicting information, they did not have enough time to suppress this task-irrelevant emotion in the literature whilst they could do so in the present research; consequently, such a negative feeling affected tasks performance in the literature but not in this study. This tentative scenario was suggested based on the study conducted by Zamorano et al. (2014). The researchers reported that as the inter-stimulus intervals became longer, inhibition occurred more robustly that enhanced response accuracy.

Whereas they focused on the duration between the stimuli, their findings could imply that having enough time when processing information would induce the stronger inhibition and increase accuracy of task performance. Combining this implication with the literature (Damen et al., 2018; Hatukai & Algom, 2017; van Steenbergen et al., 2009), it would be possible to presume the following scenario; seeing the incongruent Stroop stimuli induces negative emotions. When participants are required to respond immediately whilst they are still in a bad mood, their task performance is degraded. However, when they are allowed to spend time to overcome this task-irrelevant negative emotion, they can respond accurately with the cost of the response speed. Therefore, the present research suggests that simply feeling negative might not have robust effects on modifying ink colour effects; rather, the duration of each trial might be a key factor when considering the emotional effects in the Stroop-related tasks.

Therefore, evidence from Experiment 9, 10 and 11 could imply that spending time to process the incoming information would help learners to filter which is relevant information that enhances learning efficacy. Becoming free from the pressure on making a fast response appears to provide benefits on paired-associate learning.

As a strength, the present chapter demonstrated the way to overcome the detrimental effects from the robust semantic conflict; the negative emotions induced by the incongruent Stroop stimuli can be suppressed by extending response durations in each trial. However, the study of emotional effects with the AMST may be extended by involving the manipulation of arousal or introducing the dual-task. Since Experiment 9, 10 and 11 employed the stimuli that induced the moderate level of arousal, it could be possible to argue that the induced negative emotions from conflict consist of low pleasantness and high/low arousal; a lack of controlling the level of arousal may have led to no effects of the emotional manipulation in the present research. To explore this issue, it would be necessary to investigate at what level of arousal the incongruent Stroop stimuli induce their effects. For the research design issue, Experiment 11 illustrated that inserting the emotional stimuli between the word-number pairs would increase the risk of eliminating ink colour effects with the AMST. Therefore, when adding the emotional stimuli, this needs to be performed without modifying the inter-stimulus intervals during learning. To overcome the first limitation (i.e. a lack of arousal dimension), it would be important to manipulate the level of arousal from low to high and investigate which combination (e.g. low pleasant and high arousal) modifies task performance on the AMST. The second issue (i.e. when presenting the additional emotional stimuli) might be solved by designing the task as a dual-task. By presenting the emotional stimuli whilst learners remember the word-number pairs, it will prevent from providing the additional time to suppress the task-irrelevant emotions. Since further investigation of emotional effects on ink colour effects during the AMST may require more complex manipulations, these remain open for future research.

Overall, Experiment 9, 10 and 11 found that manipulating the level of pleasure had no notable impact on modifying the robustness of interference with the AMST. Since the duration of processing information seems to be a key factor when considering ink colour effects on the AMST, the account that memory impairment in Chapter 2 occurred due to negative emotions would be not convincing; the extension of the duration in each trial might be considered a more influential factor than simply feeling positive or negative during study whilst the possibility that emotions affect memory performance on the AMST still remains open.

Chapter 5: The Effects of Study Duration and Semantic Relationships in the AMST

Throughout 12 experiments, the present research illustrated a similar finding of ink colour effects with colour names whilst disappearance of such effects with colour-related words contradicts the original studies (Hazan-Liran & Miller, 2017; P. Miller et al., 2018). Since the former outcome was consistent with Hazan-Liran and colleagues' studies, the following discussion focuses on the latter result. The possible account for the latter result would be that a longer study duration might weaken the ink colour effects with colour-related words. Specifically, it would be possible that the amounts of extraneous cognitive load in the present research were smaller than the original studies. Consequently, ink colour difference with colour-related words became too small to be detected. The contradicting findings with colour-related words between the original studies (Hazan-Liran & Miller, 2017; P. Miller et al., 2018) and the present research might be explained by applying the formula used to calculate cognitive load with considering semantic relatedness. Oberauer and Lewandowsky (2011) introduced the formula to calculate the amount of cognitive load that was based on the Time-Based-Resource-Sharing (TBRS) theory (Barrouillet, Bernardin, & Camos, 2004).

$$\text{Cognitive Load} = \frac{\text{The duration for processing each step} \times \text{The number of processing steps}}{\text{The total time for completing a task}}$$

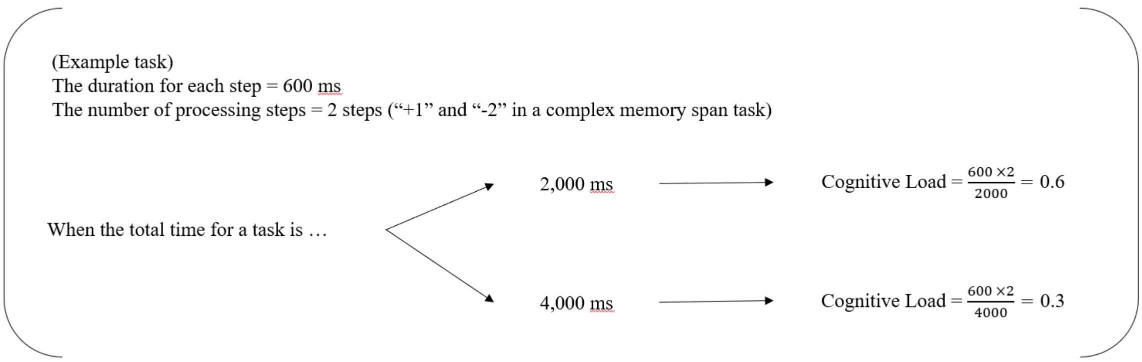


Figure 5.1. The formula to calculate the amounts of cognitive load and the example with using this formula (Oberauer & Lewandowsky, 2011).

This formula could imply that as a longer time is provided to do something, a learner would feel easier to do a task. Applying this implication with considering a semantic relationship factor would explain why the different study duration across the two original studies and the current research might have contributed to different results with colour-related words. In the present research the participants had a longer time to learn each word-number pair than the original studies. This could suggest that such learning might have been less cognitively demanding in the current study.

The formula introduced by Oberauer & Lewandovsky (2011) provided mathematical support for this explanation. Since the original studies and the present studies employed identical word-number pairs, the duration of processing each mental step and the number of such steps would be assumed to be the same. Thus, the only difference occurred in the total time for learning each word-number pair. In the two original research the study duration for each word-number pair could be calculated by the number of completion in 2 minutes. For example, Table 1 in Hazan-Liran and Miller (2017) indicated that the participants learned and wrote down the digits in less than or approximately 2 seconds in the congruent condition (e.g. 120 seconds / 86.23 completions = 1.39 seconds). In the present experiment the participants studied each word-number pair for 3 seconds. Thus, when studying materials in the AMST, the imposed amounts of cognitive load with each ink colour type became smaller than the original studies. For example, it was easier to learn word-number pairs with incongruent ink colours in the present research compared to Hazan-Liran and colleagues' research (see Table 5). Since the difference between each colour congruency condition became smaller, a longer study duration weakened the ink colour effects in the present research. This might explain the different pattern of the results between the original studies and the present studies. However, the issue of this formula was the inability of explaining the different results between colour names and colour-related words in the current study. Since the study duration was the same across these word types, consideration of only study duration would be difficult to explain the results.

One potential factor would be the number of semantic neighbours (for a summary see Levin & Tzelgov, 2016). The basic assumption of this semantic neighbour account was that when semantic information is activated, this also triggers related concepts that share the same feature or co-occur frequently in the language (for an overview see Buchanan, Westbury, & Burgess, 2001). With the assumption that a higher number of semantic neighbours produces weaker activation to the related concepts (Buchanan et al., 2001; Collins & Loftus, 1975; Levin & Tzelgov, 2016), it is expected that the number of these semantic neighbours is higher with colour-related words than with colour names, which results in the weaker robustness of the Stroop effect with colour-related words. Since the main assumption of cognitive load theory was that inhibiting task-irrelevant information requires additional cognitive resources, inhibition of such task-irrelevant concepts would affect performance on the AMST. That is, as the words coactivate more task-irrelevant concepts, more inhibition is required; consequently, the less amounts of available cognitive resources are left for paired-associate learning. When considering the number of task-irrelevant semantic neighbours, the important point here would be that the AMST involves colour-word congruency. This feature of the AMST could suggest that not all semantic neighbours are task irrelevant when using this paradigm. For instance, in the congruent condition seeing the word *sky* and the activation of the semantic neighbour *blue* could help learners to remember this word and the paired digit. Thus, the participants did not need to inhibit all semantic neighbours when learning items in the AMST. They needed to inhibit semantic neighbours that are out of colour-word links. The present thesis named these task-irrelevant semantic neighbours as semantic non-associative neighbours (SNN) and this factor would also contribute to the amounts of extraneous cognitive load in the AMST. Therefore, the present research would suggest to add this factor to the formula (see Figure 5.2).

$$\text{Cognitive Load} = \frac{\text{The duration for processing each step} \times \text{The number of processing steps}}{\text{The total time for completing a task} + \text{The number of semantic non-associative neighbours}}$$

Figure 5.2. The tentative formula to calculate the amounts of extraneous cognitive load when using the AMST.

Therefore, the tentative formula would offer two important factors when using the AMST: study duration and SNN. That is, when a study duration is long enough and learning materials have more SNNs, more inhibition of such task-irrelevant information is required. As inhibition uses up the limited cognitive resources, the less amounts of such resources are left for processing task-irrelevant colour-word associations. Consequently, task-irrelevant ink colour effects on the AMST become weaker; learners can learn materials on the AMST identically accurate regardless of contextual ink colours. Therefore, the present thesis would suggest that when the duration of learning and the number of SNN reach at certain points, ink colour effects become no longer influential on the quality of paired-associate learning with the AMST (see Table 5).

Table 5

The Tentative Amounts of Extraneous Cognitive Load in Each Colour Congruency Condition Comparing Hazan-Liran and colleagues' (Hazan-Liran & Miller, 2017; P. Miller et al., 2018) Studies and The Present Thesis

		Cognitive Load = $t_a N/ (T + SNN)$ (t_a = a duration of processing each step) (N = the number of processing steps) (T = the total time to complete a task) (SNN = the number of semantic non-associative neighbours)	
		Colour Names	Colour-related Words
Hazan-Liran & Miller (2017)	Congruent	1 x 0 / (1.39 + 0) = 0	1 x 0 / (1.59 + 10) = 0
	Incongruent	1 x 3 / (2.08 + 0) = 1.44	1 x 3 / (1.88 + 10) = 0.25
	Neutral	1 x 1 / (1.70 + 0) = 0.59	1 x 1 / (1.71 + 10) = 0.09

P. Miller et al. (2018)	Incongruent	1 x 3 / (2.08 + 0) = 1.44	1 x 3 / (1.88 + 10) = 0.25
	Neutral	1 x 1 / (1.70 + 0) = 0.59	1 x 1 / (1.71 + 10) = 0.09
The Present Thesis	Congruent	1 x 0 / (3.00 + 0) = 0	1 x 0 / (3.00 + 10) = 0
	Incongruent	1 x 3 / (3.00 + 0) = 1.00	1 x 3 / (3.00 + 10) = 0.23
	Neutral	1 x 1 / (3.00 + 0) = 0.33	1 x 1 / (3.00 + 10) = 0.08

Note. In the column of colour-related words the number of SNNs was when studying the word *sky* in the AMST. Based on Appendix C in the University of South Florida Free Association, Rhyme, and Word Fragment Norms (Douglas L. Nelson, McEvoy, & Schreiber, 2004), this word may activate 11 semantic neighbours: *blue*, *cloud*, *sun*, *heaven*, *air*, *high*, *ground*, *plane*, *space*, *star* and *up*. Since, except *blue*, other 10 words would be completely irrelevant to the AMST. Thus, the word *sky* was assumed to have 10 SNNs.

Such a suggestion for criterion of ink colour effects would be a strength of offering a mathematical formula. In Hazan-Liran and colleagues' (Hazan-Liran & Miller, 2017; P. Miller et al., 2018) reports the researchers explained that since the amounts of extraneous cognitive load were different between each ink colour congruency condition, ink colour effects were observed. However, with their verbal explanation, it would be still unclear in what circumstance such a difference becomes large enough to produce effects. In other words, the applicability of their findings in other situations would be questionable. By providing the mathematical formula to calculate such a difference, it would extend the applicability of the findings regarding ink colour effects in the AMST. Therefore, further investigation of this tentative formula will provide additional support for extending the applicability of the AMST.

Chapter 6: General Discussion

The present thesis explored the question of whether the Associative Memory Stroop Task (AMST) can be a paradigm to measure the effects of task-irrelevant colour-word associations on the quality of paired-associate learning. Based on Hazan-Liran and colleagues' (Hazan-Liran & Miller, 2017; P. Miller et al., 2018) studies, the current research aimed to explore whether the facilitative and interference effects can be observed by measuring ink colour effects on the AMST through memory accuracy on cued-recall and recognition tests. Therefore, the hypothesis was that when ink colours printed in numbers were congruent with semantics of paired colour words, this would facilitate memory accuracy on both memory tests compared to the neutral ink colour condition (i.e. facilitation) whilst when these were incongruent, it would interfere with memory accuracy compared to the neutral ink colour condition (i.e. interference). This hypothesis was tested by combining the AMST with cued-recall memory tests and associative recognition tests. The participants were asked to remember the pairs of colour concepts and numbers (e.g. *red-7*) on the AMST for a later memory test. During this task, the colour concepts were written in a neutral ink colour whilst the paired numbers were presented in other ink colours. In the congruent condition the numbers were printed in congruent ink colours (e.g. the number *7* was coloured in red ink) and these were written in incongruent ink colours in the incongruent condition (e.g. *7* printed in blue ink). In the neutral condition the numbers were presented in a neutral ink colour such as black ink or white ink. After the participants studied these concept-number pairs, they were asked to recall the paired numbers when the colour concept was used as a cue (i.e. cued-recall tests) or to judge whether the presented pairs were correct pairings or incorrect ones (i.e. associative recognition tests; Arndt, 2012). Note that, on the associative recognition tests, both colour concepts and numbers were presented in a neutral ink colour during the memory tests. The number of correct responses on these memory tests was mainly analysed. Overall, the hypothesis was partially supported; the ink colour effects were observed when applying the AMST to memory research but only with colour names.

Chapter 2 explored whether the findings of Hazan-Liran and colleagues' (Hazan-Liran & Miller, 2017; P. Miller et al., 2018) studies would be replicated when the effects of ink colours on the AMST were measured by traditional memory tests. Thus, this chapter aimed to confirm the validity of this paradigm as a tool to study learning efficacy by extending the effects reported from these two original studies. Since both facilitation and interference was observed in the original studies, it was hypothesised that similar effects would be observed with colour names and colour-related words on cued-recall tests and associative recognition tests. To achieve this aim, the participants were asked to remember the pairs of colour names or colour-related words and numbers for the later memory tests. The findings of Chapter 2 provided partial support for the hypothesis. In Experiment 1 the participants were asked to remember the pairs of colour names and numbers (e.g. *blue-4*) on the AMST and these memories were tested on the subsequent cued-recall tests. The analyses revealed both facilitation and interference whilst facilitation seemed to occur when the participants became familiar with the task. Experiment 2 also employed the cued-recall tests but the words on the AMST were colour-related words (e.g. *sky*). Differing from Experiment 1, this demonstrated no ink colour effects; the participants' memory accuracy was identical regardless of the ink colours printed in the numbers. In Experiment 3 the associative recognition tests were used to study the effects with colour names and this experiment demonstrated the poorest recognition memory accuracy in the incongruent condition. Experiment 4 studied the same tests with colour-related words and the analyses again revealed no ink colour effects on recognition memory accuracy. Combining these, whilst there was no difference on a reaction time dimension between the three colour congruency conditions, memory accuracy was affected by colour congruency when the words on the

AMST were colour names. Additionally, as the effects were observed on both memory tests, it would be possible to conclude that both types of tests can reflect ink colour effects on memory performance with the AMST. The observed interference effects could be explained by the cognitive load account (for a summary see Hazan-Liran & Miller, 2017). That is, due to suppression of task-irrelevant information, the amount of available cognitive resources was reduced for remembering the pairs in the incongruent condition. No facilitation occurred on the associative recognition tests, which might be approached from the perspective of the encoding specificity principle (Tulving & Thomson, 1973). Since the ink colours printed in the pairs were the same at studying and testing in the neutral condition on this type of memory tests, this might have enhanced memory accuracy in this condition; difference between the congruent and the neutral conditions became smaller than on the cued-recall tests. Consequently, facilitation was observed only on the cued-recall tests. Since cognitive load theory seemed to be insufficient to explain the results with colour-related words, Chapter 3 investigated this issue.

Chapter 3 examined what factor(s) the existing models of cognitive load theory need to include by exploring whether there is a situation that colour-related words produce effects with the AMST in memory research. Thus, the aim of this chapter was to study whether word concreteness, sleep consolidation processes, priming and the stimulus presentation format modify the robustness of ink colour effects with colour-related words when such effects are measured by associative recognition tests. Through these manipulations, it was hypothesised that interference would be observed with colour-related words as this effect consistently appeared in the previous chapter. This hypothesis was tested by using emotional words, having sleep before testing, pre-activation of colour-word associations and presenting colour concepts in a pictorial format. However, the analyses of these experiments demonstrated that none of these factors could modify the robustness of ink colour effects. Experiment 5-A and

5-B explored whether decreasing word concreteness could induce effects from colour-related words with the AMST. Since it was uncertain which type of emotional words (i.e. positive and negative) would be appropriate, Experiment 5-A aimed to determine this. Based on the literature that studied emotional words and memory (e.g. Pierce & Kensinger, 2011), negative emotional words were expected to show effects more likely than positive emotional words. In this experiment the participant studied the pairs of colour names and emotional words (i.e. positive, negative or neutral) for the later associative recognition tests. To reduce the complexity of the research design, colour congruency was not manipulated in Experiment 5-A; both words were presented in a neutral ink colour. Since the analyses revealed the poorest memory accuracy with the negative emotional words, the next experiment used this type of abstract colour-related words. The aim of Experiment 5-B was to examine whether congruency between task-irrelevant ink colours printed in numbers and the semantics of the paired negative emotional words on the AMST affect recognition memory accuracy. Therefore, the hypothesis of Experiment 5-B was that due to increasing difficulty of associating abstract words and numbers printed in incongruent ink colours (Paivio, 1965, 1991, 2006; Paivio & Csapo, 1969; Paivio et al., 1966; Yarmey & Paivio, 1965), interference would be observed. To test this hypothesis, the participants were instructed to study the pairs of negative emotional words and numbers (e.g. *fear-1*) with the ink colour manipulation. Despite the manipulation of word concreteness, this experiment revealed no ink colour effects with the abstract colour-related words on memory accuracy, which may have suggested no involvement of this factor in the present research.

Experiment 6 examined whether transferring short-lived memories into a long-term format would be necessary to induce effects with colour-related words on the AMST in memory research. Therefore, the aim of this experiment was to assess whether having consolidation processes via sleep could modify the robustness of ink colour effects with colour-related words. It was hypothesised that the pairs that underwent sleep would reveal the interference effect. The participants studied different pairs of colour-related words and numbers on the first day (e.g. *sky-1*) and the second day (e.g. *sapphire-8*). The memories for the two sets of the word-number pairs were tested on the second day; thus, only the Day 1 items were expected to undergo the sleep consolidation processes. However, the analyses again revealed no ink colour effects regardless of sleep experience; having sleep or not may not be a factor for no ink colour effects with colour-related words in the present research. Experiment 7 explored whether pre-activation of colour-word associations could modify the robustness of ink colour effects with colour-related words. Since the literature that studied the Stroop effect succeeded in modifying the strength of the effects by priming (e.g. Levin & Tzelgov, 2016), Experiment 7 aimed to examine whether experiencing a priming task that involved judgment based on colour-word relationships would induce effects from colour-related words. Thus, the hypothesis was that those who underwent such a priming task would demonstrate interference on the subsequent associative recognition tests. To test this hypothesis, half of the participants experienced a semantic priming task (i.e. a task that asked them to judge whether the presented colour-related words and coloured patches were related or not) whilst half of them underwent a lexical priming task (i.e. a task that required them to judge whether the presented alphabets were included in the paired colour-related words) before the main task. Note that the colour-related words during these priming tasks were different from the ones in the main task. Despite the hypothesis, this experiment again revealed no ink colour effects with both types of priming tasks, which might have suggested that pre-activation of colour-word associations was irrelevant to the current research. As a final experiment in Chapter 3, Experiment 8 tested whether changes in the stimulus presentation format would modify the robustness of ink colour effects when the AMST was combined with the associative recognition tests. The aim of this experiment was to examine whether seeing the pictures of colour-related concepts would induce any effects on memory accuracy whilst written colour-related words replicated no ink colour effects as previous experiments. Therefore, it was hypothesised that when presenting colour-related concepts in a pictorial format, interference would be observed. To test this, half of the participants studied the pairs of monochrome pictures of colour-related concepts and numbers (e.g. a black-and-white picture of a pig and the number *4*) whilst half of them studied the pairs of colour-related words and numbers (e.g. *pig-4*) and these memories were tested on the later memory tests. However, the analyses demonstrated no ink colour effects with both types of stimuli; types of the stimulus presentation formats seemed to be not involved in the present research.

Across five experiments, none of the suggested factors could induce effects from colour-related words when such effects were measured by the associative recognition memory tests. The failure of observing effects with these factors (which are common factors to be manipulated in memory research) would suggest to approach difference between colour names and colour-related words based on the Stroop literature rather than the memory studies. That is, semantic relationships between ink colours and colour-related words (for a summary see Levin & Tzelgov, 2016) might be weak enough to be neutralised by the extended duration of study. Compared to the original studies (Hazan-Liran & Miller, 2017; P. Miller et al., 2018), the duration of study was longer in the present research. Since a longer available time for performing a task reduces the amount of cognitive load (for an overview see Oberauer & Lewandowsky, 2011), this might have neutralised the detrimental effect with colour-related words whilst such effects with colour names were still valid with this longer duration of study. Although evidence from Chapter 3 suggested that duration of study and semantic relatedness might be potential factors that the existing models of cognitive load need to include, the effects of another factor with the AMST was unknown; emotional factors.

Chapter 4 investigated the effects of negative emotions triggered by conflict on memory performance with the AMST. The basic assumptions of this chapter were that processing task-irrelevant emotional information requires additional cognitive resources (for an overview see Plass & Kalyuga, 2019) and that incongruency between ink colours and colour words induce negative emotions (Botvinick et al., 2001; Dreisbach & Fischer, 2012; Hatukai & Algom, 2017; Schouppe et al., 2012; van Steenbergen et al., 2009). Combining these, Chapter 4 aimed to examine whether suppression of task-irrelevant negative emotions due to colour-word incongruency produced memory impairment when ink colour effects on the AMST were measured by the cued-recall tests. Therefore, the hypothesis was that increasing unpleasantness of learning materials and/or elevating learners' unpleasant feeling would mitigate the interference effect due to mood congruency (V. E. Lewis & Williams, 1989; Plass & Kalyuga, 2019). To test this hypothesis, the good and bad numbers, the face icons and the emotional sounds were presented during study and the memories of the word-number pairs were tested on the cued-recall tests. Despite these manipulations, the robustness of ink colour effects was identical; the hypothesis was not supported. Experiment 9 explored a question of whether unavoidable conflict due to incongruent ink colours led to memory impairment and if so, whether manipulation of pleasure with a moderate level of arousal modified the robustness of ink colour effects with the AMST. To test this, the aim of Experiment 9 was to examine whether assigning the good and bad values to the numbers modified cued-recall memory accuracy with the AMST. Since the literature (Hatukai & Algom, 2017) reported the mitigation of the Stroop interference effect by employing the "good" and "bad" responses, the hypothesis was that studying the bad numbers would enhance memory accuracy than studying the good numbers in the incongruent condition. Therefore, the good and bad values were assigned to each number whilst the participants studied the pairs of colour names and these numbers on the AMST for the later cued-recall memory tests. The analyses, however, revealed no changes in the robustness of effects; facilitation and interference occurred with both types of numbers. Experiment 10 tested a question of whether interference caused by negative emotions could be mitigated by exposure to positive-valenced stimuli if such an emotional factor affects memory performance with the AMST. The aim of this experiment was to assess whether seeing the three types of face icons (i.e. happy, sad and neutral) during study would modify the robustness of ink colour effects on the AMST. Based on the literature (van Steenbergen et al., 2009) that found the mitigation of the Stroop interference effect by seeing the smiley face icons (which indicated monetary gain), it was hypothesised that, in the incongruent condition, seeing the happy face icons during study would reduce the robustness of interference on cued-recall memory accuracy than seeing the sad and neutral face icons. This hypothesis was tested by presenting the face icons between each word-number pair whilst the participants were told to ignore these face icons. These face icons were presented in black ink. The analyses indicated that the robustness of ink colour effects on the AMST was stable regardless of the types of face icons. Interestingly, the analyses revealed the occurrence of facilitation; no interference was observed. Whilst this experiment may provide support for the duration of study account suggested from Chapter 3, to overcome the methodological issue of Experiment 10, Experiment 11 used the three types of the emotional sounds (i.e. positive, negative and neutral) during study. These emotional stimuli were presented for a longer duration than Experiment 10 in order for the participants to recognise these sounds precisely and no response was required for these auditory stimuli. Differing from the previous experiments, Experiment 11 demonstrated no ink colour effects with the colour names on cued-recall memory accuracy, which might be explained by the longer inter-stimulus intervals during study (Zamorano et al., 2014). Evidence from Chapter 4 indicated that duration of the task might be more influential than an emotional factor on the AMST in memory research. Since the longer duration for completing a task produced the mitigation of ink colour effects on the AMST whilst manipulations on the emotional dimension resulted in little impact on modifying the robustness of such effects, this could signal that a temporal factor would be a potential factor that the present research needs to consider in addition to semantic relatedness.

To summarise, when task-irrelevant ink colour effects on the AMST were measured by cued-recall and associative recognition tests, the following were found; (a) both tests demonstrated effects, (b) interference was consistently observed on both memory tests whilst the occurrence of facilitation was less consistent than of interference; (c) these effects, however, occurred only with colour names; colour-related words did not show any ink colour effects on memory performance, (d) such effects on memory accuracy seemed to involve semantic relatedness between ink colours and colour concepts and duration of a task rather than an emotional factor during study. Since ink colour effects were observed on traditional memory tests, the present research would conclude that the findings of Hazan-Liran and colleagues (Hazan-Liran & Miller, 2017; P. Miller et al., 2018) occurred genuinely on learning efficacy and the AMST can be used as a new paradigm in research of paired-associate learning.

In addition to these main findings, the analyses of the present experiments kept revealing that the strength of ink colour effects was stable across practice. The assumption of the practice effects was retrieved from memory confusion during a retrieval process (for a summary see Wixted, 2005). Thus, the pattern of the results in the present research might signal ink colour effects at encoding than at retrieval. As the investigation of the timing of ink colour effects may extend the usefulness of the AMST, this can be something that future research can explore.

Evidence from the present research would imply that consideration of study duration prevents the harmful effect on the quality of paired-associate learning that is triggered by the semantic interaction between learning materials and to-be-ignored information.

Importantly, this was the first research that attempted to study the credibility of the AMST from the perspective of memory accuracy. The findings of the present thesis extended Hazan-Liran and colleagues' (Hazan-Liran & Miller, 2017; P. Miller et al., 2018) studies by illustrating disappearance of the harmful ink colour effects with colour-related words in the longer study time, which would suggest a new possibility that providing enough study time prevents such negative effects on learning efficacy when ink colours and colour words are indirectly associated. However, there seems to be several limitations to work in the present research. First, the limited variety of ink colours and colour-related words would suggest that using different ink colours and colour-related words might produce different results. As the pattern of the results were different between cued-recall tests and recognition tests, ink colour congruency during study and tests might modify the robustness of the effects based on the literature (Vernon, 2001) that found such effects on the memory tests. Since the present research does not exclude the possibility of emotional effects on the AMST, the investigation of whether what combination of pleasure and arousal contribute to the Stroop-related effects would also add work to the original research (Hazan-Liran & Miller, 2017; P. Miller et al., 2018) and the current study. Additionally, the suggested account with the tentative formula would raise a question of whether manipulation of the study time and semantic relatedness modify task performance on the AMST.

These limitations might be overcome in the following way. The effects of ink colours between study and tests can be examined with recognition tests. For example, presenting the test items with congruent, incongruent or neutral ink colours not only to colour words but also to the previous status (e.g. presenting in red ink at study but presenting in blue ink at test) may demonstrate evidence of how such colour congruency affects performance on the subsequent memory tests. The use of cued-recall tests would offer two potential pathways for future research: approach from emotions and approach from the tentative model of cognitive load theory. To tackle ink colour effects on the AMST from the emotional perspective, manipulation of arousal will be a key factor. That is, between three ink colour congruency conditions, it can be studied that whether high pleasure and high or low arousal would modify the magnitude of facilitation whilst low pleasure and high or low arousal would change the robustness of interference with the AMST. To investigate such ink colour effects with the tentative model of cognitive load theory, there will be two factors to be manipulated: a duration of study and a semantic non-associative neighbour. Because reduction of study time might increase the task difficulty, comparing task performance between the 3-second study time and longer times may confirm whether duration of study affects ink colour effects with the AMST. The manipulation of semantic non-associative neighbours may be achieved by two steps: examining semantic neighbours and sorting these out based on their semantic relatedness to colour words. For instance, the word *tomato* might activate the concept of red, vegetable and healthy. However, the word *apple* may trigger the concept of red, fruit and black since this word can indicate a specific electronic company logo. The number and type of such semantic neighbours can be evaluated through the word association database (e.g. the University of South Florida Free Association, Rhyme, and Word Fragment Norms; Nelson et al., 2004). Whether each semantic neighbour is related to a specific colour word can be examined by using lexical semantic models (e.g. the Latent Semantic Analysis; for an overview see Rohde, Gonnerman, & Plaut, 2006). In the case of *tomato*, it will be possible to expect that the concept of red shows a strong semantic association with a colour word *red* whilst the other two semantic neighbours (i.e. vegetable and healthy) may not; there will be two semantic non-associative neighbours. With the word *apple*, the co-activated concepts of red and black will demonstrate high semantic relatedness values with the colour words whilst the concept of fruit might not; there will be one semantic non-associative neighbour. By combining these analyses with the tentative model of cognitive load theory, it might allow researchers to investigate whether the suggested model can describe ink colour effects on the AMST.

Overall, the present research provided support for the idea that the AMST can be used as a paradigm to study learning efficacy. Specifically, the use of the cued-recall tests and recognition tests consistently demonstrated impairment of memory accuracy with incongruent ink colours printed in the numbers when the words on the AMST were the colour names whilst no ink colour effects were observed with the colour-related words. This contradicting evidence against the original studies (Hazan-Liran & Miller, 2017; P. Miller et al., 2018) might signal to include a duration of study and semantic relatedness with the concept of cognitive load. Therefore, the present research suggests that the consideration of study time and semantics would be keys to optimise learning efficacy.

References

Alvarez, G. A., & Cavanagh, P. (2004). The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychological Science*, *15*(2), 106–111. https://doi.org/10.1111/j.0963-7214.2004.01502006.x

Anderson, M. C. (2003). Rethinking interference theory: Executive control and the mechanisms of forgetting. *Journal of Memory and Language*, *49*(4), 415–445. https://doi.org/10.1016/j.jml.2003.08.006

Ankala, V. (2011). Retroactive Interference and Forgetting. *Undergraduate Journal of Mathematical Modeling: One + Two*, *3*(2). https://doi.org/10.5038/2326-3652.3.2.4

Arndt, J. (2012). Paired-associate learning. In N. M. Seel (Ed.), *Encyclopedia of the Sciences of Learning* (pp. 2551–2552). https://doi.org/10.1007/978-1-4419-1428-6

Baddeley, A. (2012). Working memory: Theories, models, and controversies. *Annual Review of Psychology*, *63*, 1–29. https://doi.org/10.1146/annurev-psych-120710-100422

Baddeley, A. D., & Hitch, G. (1993). The recency effect: Implicit learning with explicit retrieval? *Memory & Cognition*, *21*(2), 146–155. https://doi.org/10.3758/BF03202726

Barrouillet, P., Bernardin, S., & Camos, V. (2004). Time constraints and resource sharing in adults' working memory spans. *Journal of Experimental Psychology: General*, *133*(1), 83–100. https://doi.org/10.1037/0096-3445.133.1.83

BEKA - BR385 tone samples. (2017). Retrieved October 16, 2017, from BEKA associates Ltd website: https://www.beka.co.uk/br385_tone_samples.html

Bentin, S., & Moscovitch, M. (1988). The time course of repetition effects for words and unfamiliar faces. *Journal of Experimental Psychology: General*, *117*(2), 148–160. https://doi.org/10.1037/0096-3445.117.2.148

Besner, D., & Stolz, J. A. (1999). What kind of attention modulates the Stroop effect? *Psychonomic Bulletin & Review*, *6*(1), 99–104. https://doi.org/10.3758/BF03210815

Bird, C. M. (2017). The role of the hippocampus in recognition memory. *Cortex*, *93*(0), 155–165. https://doi.org/10.1016/j.cortex.2017.05.016

Bisby, J. A., & Burgess, N. (2014). Negative affect impairs associative memory but not item memory. *Learning and Memory*, *21*(1), 21–27. https://doi.org/10.1101/lm.032409.113

Bisby, J. A., Horner, A. J., Hørlyck, L. D., & Burgess, N. (2016). Opposing effects of negative emotion on amygdalar and hippocampal memory for items and associations. *Social Cognitive and Affective Neuroscience*, *11*(6), 981–990. https://doi.org/10.1093/scan/nsw028

Boksem, M. A. S., & Tops, M. (2008). Mental fatigue: Costs and benefits. *Brain Research Reviews*, *59*(1), 125–139. https://doi.org/10.1016/j.brainresrev.2008.07.001

Botvinick, M. M. (2007). Conflict monitoring and decision making: Reconciling two perspectives on anterior cingulate function. *Cognitive, Affective & Behavioral Neuroscience*, *7*(4), 356–366. https://doi.org/10.3758/CABN.7.4.356

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, *108*(3), 624–652. https://doi.org/10.1037/0033-295X.108.3.624

Bradley, M. M., & Lang, P. J. (1999). Affective norms for English words (ANEW): Instruction manual and affective ratings. In *Technical Report C-1, The Center for Research in Psychophysiology*. Retrieved from University of Florida website: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.306.3881&rep=rep1&type=pdf

Bradley, M. M., & Lang, P. J. (2007). The International Affective Digitized Sounds (2nd Edition; IADS-2): Affective ratings of sounds and instruction manual. In *(Technical Report No. B-3)*. Gainesville, FL: University of Florida.

Brainerd, C. J., Stein, L. M., Silveira, R. A., Rohenkohl, G., & Reyna, V. F. (2008). How does negative emotion cause false memories? *Psychological Science*, *19*(9), 919–925. https://doi.org/10.1111/j.1467-9280.2008.02177.x

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods*, *46*(3), 904–911. https://doi.org/10.3758/s13428-013-0403-5

Buchanan, L., Westbury, C., & Burgess, C. (2001). Characterizing semantic space: Neighborhood effects in word recognition. *Psychonomic Bulletin & Review*, *8*(3), 531–544. https://doi.org/10.3758/BF03196189

Campos, T. F., Barroso, M. T. M., & de Lara Menezes, A. A. (2010). Encoding, storage and retrieval processes of the memory and the implications for motor practice in stroke patients. *NeuroRehabilitation*, *26*(2), 135–142. https://doi.org/10.3233/NRE-2010-0545

Ceraso, J. (1967). The interference theory of forgetting. *Scientific American*, *217*(4), 117–127. https://doi.org/10.1038/scientificamerican1067-117

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, *82*(6), 407–428. https://doi.org/10.1037/0033-295X.82.6.407

Colour. (2020). In *Wikipedia*. Retrieved August 25, 2019 from https://simple.wikipedia.org/wiki/Colour

Cowan, N. (2001). Metatheory of storage capacity limits. *Behavioral and Brain Sciences*, *24*(1), 154–176. https://doi.org/10.1017/S0140525X0161392X

Cowan, N., Towse, J. N., Hamilton, Z., Saults, J. S., Elliott, E. M., Lacey, J. F., … Hitch, G. J. (2003). Children's working-memory processes: A response-timing analysis. *Journal of Experimental Psychology: General*, *132*(1), 113–132. https://doi.org/10.1037/0096-3445.132.1.113

Craik, F. I., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning & Verbal Behavior*, *11*(6), 671–684. https://doi.org/10.1016/S0022-5371(72)80001-X

D'Eredita, M. A., & Hoyer, W. J. (2010). Transfer of instances in cognitive skill learning: Adult age differences. *Experimental Aging Research*, *36*(1), 23–39. https://doi.org/10.1080/03610730903418646

Damen, T. G. E., Strick, M., Taris, T. W., & Aarts, H. (2018). When conflict influences liking : The case of the Stroop task. *PloS One*, *13*(7), 1–23. https://doi.org/10.1371/journal.pone.0199700

Davis, O. C., Geller, A. S., Rizzuto, D. S., & Kahana, M. J. (2008). Temporal associative processes revealed by intrusions in paired-associate recall. *Psychonomic Bulletin & Review*, *15*(1), 64–69. https://doi.org/10.3758/PBR.15.1.64

de Jong, T. (2010). Cognitive load theory, educational research, and instructional design: Some food for thought. *Instructional Science*, *38*(2), 105–134. https://doi.org/10.1007/s11251-009-9110-0

De Marchis, G., Rivero Expósito, M. del P., & Reales Avilés, J. M. (2013). Psychological distance and reaction time in a Stroop task. *Cognitive Processing*, *14*(4), 401–410. https://doi.org/10.1007/s10339-013-0569-x

Debue, N., & van de Leemput, C. (2014). What does germane load mean? An empirical contribution to the cognitive load theory. *Frontiers in Psychology*, *5*(1099), 1–12. https://doi.org/10.3389/fpsyg.2014.01099

Defeyter, M. A., Russo, R., & McPartlin, P. L. (2009). The picture superiority effect in recognition memory: A developmental study using the response signal procedure. *Cognitive Development*, *24*(3), 265–273. https://doi.org/10.1016/j.cogdev.2009.05.002

Diekelmann, S., & Born, J. (2010). The memory function of sleep. *Nature Reviews Neuroscience*, *11*(2), 114–126. https://doi.org/10.1038/nrn2762

Digital Synopsis. (n.d.). It’s “Wine”, not dark red - here are the correct names of all colour shades. Retrieved August 25, 2019, from Digital Synopsis website: https://digitalsynopsis.com/design/color-thesaurus-correct-names-of-shades/

Dignath, D., & Eder, A. B. (2015). Stimulus conflict triggers behavioral avoidance. *Cognitive, Affective & Behavioral Neuroscience*, *15*(4), 822–836. https://doi.org/10.3758/s13415-015-0355-6

Dreisbach, G., & Fischer, R. (2012). Conflicts as aversive signals. *Brain and Cognition*, *78*(2), 94–98. https://doi.org/10.1016/j.bandc.2011.12.003

Dreisbach, G., & Goschke, T. (2004). How positive affect modulates cognitive control: Reduced perseveration at the cost of increased distractibility. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(2), 343–353. https://doi.org/10.1037/0278-7393.30.2.343

Entel, O., Tzelgov, J., Bereby-Meyer, Y., & Shahar, N. (2015). Exploring relations between task conflict and informational conflict in the Stroop task. *Psychological Research*, *79*(6), 913–927. https://doi.org/10.1007/s00426-014-0630-0

Erk, S., Martin, S., & Walter, H. (2005). Emotional context during encoding of neutral items modulates brain activation not only during encoding but also during recognition. *NeuroImage*, *26*(3), 829–838. https://doi.org/10.1016/j.neuroimage.2005.02.045

Feenan, K., & Snodgrass, J. G. (1990). The effect of context on discrimination and bias in recognition memory for pictures and words. *Memory & Cognition*, *18*(5), 515–527. https://doi.org/10.3758/BF03198484

Fenn, K. M., & Hambrick, D. Z. (2012). Individual differences in working memory capacity predict sleep-dependent memory consolidation. *Journal of Experimental Psychology: General*, *141*(3), 404–410. https://doi.org/10.1037/a0025268

Fenn, K. M., & Hambrick, D. Z. (2013). What drives sleep-dependent memory consolidation: Greater gain or less loss? *Psychonomic Bulletin & Review*, *20*(3), 501–506. https://doi.org/10.3758/s13423-012-0366-z

Gagnepain, P., Lebreton, K., Desgranges, B., & Eustache, F. (2008). Perceptual priming enhances the creation of new episodic memories. *Consciousness and Cognition: An International Journal*, *17*(1), 276–287. https://doi.org/10.1016/j.concog.2007.03.006

Gais, S., Plihal, W., Wagner, U., & Born, J. (2000). Early sleep triggers memory for early visual discrimination skills. *Nature Neuroscience*, *3*(12), 1335–1339. https://doi.org/10.1038/81881

Goldfarb, L., & Henik, A. (2007). Evidence for task conflict in the Stroop effect. *Journal of Experimental Psychology: Human Perception and Performance*, *33*(5), 1170–1176. https://doi.org/10.1037/0096-1523.33.5.1170

Goldfarb, L., & Treisman, A. (2010). Are some features easier to bind than others? The congruency effect. *Psychological Science*, *21*(5), 676–681. https://doi.org/10.1177/0956797610365130

Golomb, J. D., Peelle, J. E., Addis, K. M., Kahana, M. J., & Wingfield, A. (2008). Effects of adult aging on utilization of temporal and semantic associations during free and serial recall. *Memory & Cognition*, *36*(5), 947–956. https://doi.org/10.3758/MC.36.5.947

Haist, F., Shimamura, A. P., & Squire, L. R. (1992). On the relationship between recall and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(4), 691–702. https://doi.org/10.1037/0278-7393.18.4.691

Hales, J. B., & Brewer, J. B. (2011). The timing of associative memory formation: Frontal lobe and anterior medial temporal lobe activity at associative binding predicts memory. *Journal of Neurophysiology*, *105*(4), 1454–1463. https://doi.org/10.1152/jn.00902.2010

Hatukai, T., & Algom, D. (2017). The Stroop incongruity effect: Congruity relationship reaches beyond the Stroop task. *Journal of Experimental Psychology: Human Perception and Performance*, *43*(6), 1098–1114. https://doi.org/10.1037/xhp0000381

Hazan-Liran, B., & Miller, P. (2017). Stroop-like effects in a new-code learning task: A cognitive load theory perspective. *The Quarterly Journal of Experimental Psychology*, *70*(9), 1878–1891. https://doi.org/10.1080/17470218.2016.1214845

Henik, A., Bugg, J. M., & Goldfarb, L. (2018). Inspired by the past and looking to the future of the Stroop effect. *Acta Psychologica*, *189*, 1–3. https://doi.org/10.1016/j.actpsy.2018.06.007

Hock, H., & Petrasek, J. (1973). Verbal interference with perceptual classification: The effect of semantic structure. *Perception & Psychophysics*, *13*(1), 116–120. https://doi.org/10.3758/BF03207245

Huettig, F., & Altmann, G. T. M. (2011). Looking at anything that is green when hearing “frog”: How object surface colour and stored object colour knowledge influence language-mediated overt attention. *The Quarterly Journal of Experimental Psychology*, *64*(1), 122–145. https://doi.org/10.1080/17470218.2010.481474

Huntjens, R. J. C., Postma, A., Hamaker, E. L., Woertman, L., Van Der Hart, O., & Peters, M. (2002). Perceptual and conceptual priming in patients with dissociative identity disorder. *Memory & Cognition*, *30*(7), 1033–1043. https://doi.org/10.3758/BF03194321

Ingham, J. G. (1970). Individual differences in signal detection. *Acta Psychologica, Amsterdam*, *34*(1), 39–50. https://doi.org/10.1016/0001-6918(70)90003-X

Intraub, H. (1980). Presentation rate and the representation of briefly glimpsed pictures in memory. *Journal of Experimental Psychology: Human Learning and Memory*, *6*(1), 1–12. https://doi.org/10.1037/0278-7393.6.1.1

Jodo, E., & Kayama, Y. (1992). Relation of a negative ERP component to response inhibition in a Go/No-go task. *Electroencephalography & Clinical Neurophysiology*, *82*(6), 477–482. https://doi.org/10.1016/0013-4694(92)90054-L

Johansson, S., & Hofland, K. (1989). *Frequency analysis of English vocabulary and grammar: Based on the LOB corpus, volume 1: Tag frequencies and word frequencies*. Oxford: Clarendon Press.

Joseph, J. E., & Proffitt, D. R. (1996). Semantic versus perceptual influences of color in object recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(2), 407–429. https://doi.org/10.1037/0278-7393.22.2.407

Kalyuga, S. (2007). Expertise reversal effect and its implications for learner-tailored instruction. *Educational Psychology Review*, *19*(4), 509–539. https://doi.org/10.1007/s10648-007-9054-3

Kim, J. J., & Diamond, D. M. (2002). The stressed hippocampus, synaptic plasticity and lost memories. *Nature Reviews Neuroscience*, *3*(6), 453–462. https://doi.org/10.1038/nrn849

Kiyonaga, A., & Egner, T. (2014). The working memory Stroop effect: When internal representations clash with external stimuli. *Psychological Science*, *25*(8), 1619–1629. https://doi.org/10.1177/0956797614536739

Klinzing, J. G., Niethard, N., & Born, J. (2019). Mechanisms of systems memory consolidation during sleep. *Nature Neuroscience*, *22*, 1598–1610. https://doi.org/10.1038/s41593-019-0467-3

Knickerbocker, H., & Altarriba, J. (2013). Differential repetition blindness with emotion and emotion-laden word types. *Visual Cognition*, *21*(5), 599–627. https://doi.org/10.1080/13506285.2013.815297

Kool, W., McGuire, J. T., Rosen, Z. B., & Botvinick, M. M. (2010). Decision making and the avoidance of cognitive demand. *Journal of Experimental Psychology: General*, *139*(4), 665–682. https://doi.org/10.1037/a0020198

Korman, M., Doyon, J., Doljansky, J., Carrier, J., Dagan, Y., & Karni, A. (2007). Daytime sleep condenses the time course of motor memory consolidation. *Nature Neuroscience*, *10*(9), 1206–1213. https://doi.org/10.1038/nn1959

Kousta, S. T., Vigliocco, G., Vinson, D. P., Andrews, M., & Del Campo, E. (2011). The representation of abstract words: Why emotion matters. *Journal of Experimental Psychology: General*, *140*(1), 14–34. https://doi.org/10.1037/a0021446

Laham, D., & Steinhart, D. (1998). Latent Semantic Analysis @ CU Boulder. Retrieved October 8, 2018, from University of Colorado at Boulder website: http://lsa.colorado.edu/

Lahl, O., Wispel, C., Willigens, B., & Pietrowsky, R. (2008). An ultra short episode of sleep is sufficient to promote declarative memory performance. *Journal of Sleep Research*, *17*(1), 3–10. https://doi.org/10.1111/j.1365-2869.2008.00622.x

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, *104*(2), 211–240. https://doi.org/10.1037/0033-295X.104.2.211

Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Processes*, *25*(2–3), 259–284. https://doi.org/10.1080/01638539809545028

Leech, G., Rayson, P., & Wilson, A. (2001). *Word frequencies in written and spoken English: Based on the British National Corpus*. London: Longman.

Levin, Y., & Tzelgov, J. (2016). What Klein's "semantic gradient" does and does not really show: Decomposing stroop interference into task and informational conflict components. *Frontiers in Psychology*, *7*, 249. https://doi.org/10.3389/fpsyg.2016.00249

Lewis, P. A., & Durrant, S. J. (2011). Overlapping memory replay during sleep builds cognitive schemata. *Trends in Cognitive Sciences*, *15*(8), 343–351. https://doi.org/10.1016/j.tics.2011.06.004

Lewis, V. E., & Williams, R. N. (1989). Mood-congruent vs. mood-state-dependent learning: Implications for a view of emotion. *Journal of Social Behavior & Personality*, *4*(2), 157–171. Retrieved from https://search.proquest.com/docview/1292364460/fulltextPDF/205100C856184642PQ/1?accountid=7408

Lucas, M. (2000). Semantic priming without association: A meta-analytic review. *Psychonomic Bulletin & Review*, *7*(4), 618–630. https://doi.org/10.3758/BF03212999

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin*, *109*(2), 163–203. https://doi.org/10.1037/0033-2909.109.2.163

MacLeod, C. M. (1996). How priming affects two speeded implicit tests of remembering: Naming colors versus reading words. *Consciousness and Cognition: An International Journal*, *5*(1–2), 73–90. https://doi.org/10.1006/ccog.1996.0005

MacLeod, C. M., & MacDonald, P. A. (2000). Interdimensional interference in the Stroop effect: Uncovering the cognitive and neural anatomy of attention. *Trends in Cognitive Sciences*, *4*(10), 383–391. https://doi.org/10.1016/S1364-6613(00)01530-8

Macmillan, N. A., & Creelman, C. D. (1991). *Detection theory: A user's guide*. New York, NY: Cambridge University Press.

Macmillan, N. A., & Creelman, C. D. (2005). Detection theory: A user's guide. In *Lawrence Erlbaum Associates Publishers* (2nd ed.). Retrieved from http://digitus.itk.ppke.hu/~banko/VisionGroup/SignalDetectionTheory.pdf

Madan, C. R., Scott, S. M. E., & Kensinger, E. A. (2019). Positive emotion enhances association-memory. *Emotion*, *19*(4), 733–740. https://doi.org/10.1037/emo0000465

Masson, M. E. J., & Loftus, G. R. (2003). Using confidence intervals for graphically based data interpretation. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, *57*(3), 203–220. https://doi.org/10.1037/h0087426

Mather, M. (2007). Emotional arousal and memory binding: An object-based framework. *Perspectives on Psychological Science*, *2*(1), 33–52. https://doi.org/10.1111/j.1745-6916.2007.00028.x

Matzen, L. E., Trumbo, M. C., Leach, R. C., & Leshikar, E. D. (2015). Effects of non-invasive brain stimulation on associative memory. *Brain Research*, *1624*, 286–296. https://doi.org/10.1016/j.brainres.2015.07.036

McBride, D. M., & Dosher, B. A. (2002). A comparison of conscious and automatic memory processes for picture and word stimuli: A process dissociation analysis. *Consciousness and Cognition: An International Journal*, *11*(3), 423–460. https://doi.org/10.1016/S1053-8100(02)00007-7

McClain, L. (1983). Color priming affects Stroop interference. *Perceptual and Motor Skills*, *56*(2), 643–651. https://doi.org/10.2466/pms.1983.56.2.643

MediaCollege.com - Download Audio Tone Files. (n.d.). Retrieved October 16, 2017, from MediaCollege.com website: https://www.mediacollege.com/audio/tone/download/

Medler, D. A., & Binder, J. R. (2005). MCWord: An On-Line Orthographic Database of the English Language. Retrieved October 15, 2018, from http://www.neuro.mcw.edu/mcword/

Mednick, S., Nakayama, K., & Stickgold, R. (2003). Sleep-dependent learning: A nap is as good as a night. *Nature Neuroscience*, *6*(7), 697–698. https://doi.org/10.1038/nn1078

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*(2), 81–97. https://doi.org/10.1037/h0043158

Miller, P., Hazan-Liran, B., & Cohen, D. (2018). Does task-irrelevant colour information create extraneous cognitive load? Evidence from a learning task. *Quarterly Journal of Experimental Psychology*, *72*(5), 1155–1163. https://doi.org/10.1177/1747021818781425

Mills, C. B., Innis, J., Westendorf, T., Owsianiecki, L., & McDonald, A. (2006). Effect of a synesthete's photisms on name recall. *Cortex: A Journal Devoted to the Study of the Nervous System and Behavior*, *42*(2), 155–163. https://doi.org/10.1016/S0010-9452(08)70340-X

Nakazawa, K., Quirk, M. C., Chitwood, R. A., Watanabe, M., Yeckel, M. F., Sun, L. D., … Tonegawa, S. (2002). Requirement for hippocampal CA3 NMDA receptors in associative memory recall. *Science*, *297*(5579), 211–218. https://doi.org/10.1126/science.1071795

Naor-Raz, G., Tarr, M. J., & Kersten, D. (2003). Is color an intrinsic property of object representation? *Perception*, *32*(6), 667–680. https://doi.org/10.1068/p5050

Nelson, D. L., Reed, V. S., & Walling, J. R. (1976). Pictorial superiority effect. *Journal of Experimental Psychology. Human Learning and Memory*, *2*(5), 523–528. https://doi.org/10.1037/0278-7393.2.5.523

Nelson, Douglas L., McEvoy, C. L., & Schreiber, T. A. (2004). The University of South Florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments & Computers*, *36*(3), 402–407. https://doi.org/10.3758/BF03195588

Nieuwenhuis-Mark, R. E. (2012). Recall and the effect of repetition on recall. In N. M. Seel (Ed.), *Encyclopedia of the Sciences of Learning* (pp. 2779–2782). https://doi.org/10.1007/978-1-4419-1428-6

Nishida, M., & Walker, M. P. (2007). Daytime naps, motor memory consolidation and regionally specific sleep spindles. *PLOS ONE*, *2*(4). https://doi.org/10.1371/journal.pone.0000341

Oberauer, K., & Lewandowsky, S. (2011). Modeling working memory: A computational implementation of the Time-Based Resource-Sharing theory. *Psychonomic Bulletin & Review*, *18*(1), 10–45. https://doi.org/10.3758/s13423-010-0020-6

Paivio, A. (1965). Abstractness, imagery, and meaningfulness in paired-associate learning. *Journal of Verbal Learning & Verbal Behavior*, *4*(1), 32–38. https://doi.org/10.1016/S0022-5371(65)80064-0

Paivio, A. (1991). Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology/Revue Canadienne de Psychologie*, *45*(3), 255–287. https://doi.org/10.1037/h0084295

Paivio, A. (2006). Dual coding theory and education. In *Draft chapter presented at the conference on "Pathways to literacy achievement for high poverty children," The University of Michigan School of Education, September 29–October 1, 2006*. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.329.7319&rep=rep1&type=pdf

Paivio, A. (2013). Dual coding theory, word abstractness, and emotion: A critical review of Kousta et al. (2011). *Journal of Experimental Psychology: General*, *142*(1), 282–287. https://doi.org/10.1037/a0027004

Paivio, A., & Csapo, K. (1969). Concrete image and verbal memory codes. *Journal of Experimental Psychology*, *80*(2, Pt.1), 279–285. https://doi.org/10.1037/h0027273

Paivio, A., & Yarmey, A. D. (1966). Pictures versus words as stimuli and responses in paired-associate learning. *Psychonomic Science*, *5*(6), 235–236. https://doi.org/10.3758/BF03328369

Paivio, A., Yuille, J. C., & Smythe, P. C. (1966). Stimulus and response abstractness, imagery, and meaningfulness, and reported mediators in paired-associate learning. *Canadian Journal of Psychology/Revue Canadienne de Psychologie*, *20*(4), 362–377. https://doi.org/10.1037/h0082949

Peeck, J. (1974). Retention of pictorial and verbal content of a text with illustrations. *Journal of Educational Psychology*, *66*(6), 880–888. https://doi.org/10.1037/h0021531

Peirce, J. W., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., … Lindeløv, J. K. (2019). PsychoPy2: experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Pekrun, R., Goetz, T., Titz, W., & Perry, R. P. (2002). Academic emotions in students' self-regulated learning and achievement: A program of qualitative and quantitative research. *Educational Psychologist*, *37*(2), 91–105. https://doi.org/10.1207/S15326985EP3702_4

Peterson, L. R., & Peterson, M. J. (1959). Short-term retention of individual verbal items. *Journal of Experimental Psychology*, *58*(3), 193–198. https://doi.org/10.1037/h0049234

Pierce, B. H., & Kensinger, E. A. (2011). Effects of emotion on associative recognition: Valence and retention interval matter. *Emotion*, *11*(1), 139–144. https://doi.org/10.1037/a0021287

Plass, J. L., & Kalyuga, S. (2019). Four ways of considering emotion in cognitive load theory. *Educational Psychology Review*, *31*, 339–359. https://doi.org/10.1007/s10648-019-09473-5

Plass, J. L., & Kaplan, U. (2016). Emotional design in digital media for learning. In S. Y. Tettegah & M. Gartmeier (Eds.), *Emotions, Technology, Design, and Learning* (pp. 131–161). https://doi.org/10.1016/B978-0-12-801856-9.00007-4

Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology, 32*(1), 3–25. https://doi.org/10.1080/00335558008248231

Potkin, K. T., & Bunney, W. E. (2012). Sleep improves memory: The effect of sleep on long term memory in early adolescence. *PLoS ONE*, *7*(8), :e42191. https://doi.org/10.1371/journal.pone.0042191

Pratto, F., & John, O. P. (1991). Automatic vigilance: The attention-grabbing power of negative social information. *Journal of Personality and Social Psychology, 61*(3), 380–391. https://doi.org/10.1037/0022-3514.61.3.380

Rajji, T., Chapman, D., Eichenbaum, H., & Greene, R. (2006). The role of CA3 hippocampal NMDA receptors in paired associate learning. *The Journal of Neuroscience*, *26*(3), 908–915. https://doi.org/10.1523/JNEUROSCI.4194-05.2006

Rammsayer, T. H., & Verner, M. (2015). Larger visual stimuli are perceived to last longer from time to time: The internal clock is not affected by nontemporal visual stimulus size. *Journal of Vision*, *15*(3), 1–11. https://doi.org/10.1167/15.3.5

Randomness and Integrity Services Ltd. (n.d.). RANDOM.ORG - List Randomizer. Retrieved November 10, 2017, from https://www.random.org/lists/

Rasamimanana, M., Barbaroux, M., Colé, P., & Besson, M. (2020). Semantic compensation and novel word learning in university students with dyslexia. *Neuropsychologia*, *139*(2). https://doi.org/10.1016/j.neuropsychologia.2020.107358

Reber, A. S. (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General*, *118*(3), 219–235. https://doi.org/10.1037/0096-3445.118.3.219

Reisberg, D., & Heuer, F. (2004). Memory for emotional events. In D. Reisberg & P. Hertel (Eds.), *Memory and emotion* (pp. 3–41). https://doi.org/10.1093/acprof:oso/9780195158564.001.0001

Richardson-Klavehn, A., & Bjork, R. A. (1988). Measures of Memory. *Annual Review of Psychology*, *39*(1), 475–543. https://doi.org/10.1146/annurev.ps.39.020188.002355

Rohde, D. L. T., Gonnerman, L. M., & Plaut, D. C. (2006). An improved model of semantic similarity based on lexical co-occurrence. *Communications of the ACM, 8*, 627–633. Retrieved from http://www.cnbc.cmu.edu/~plaut/papers/pdf/RohdeGonnermanPlautSUB-CogSci.COALS.pdf#:~:text=The HAL method for modeling semantic memory %28Lund,the similarity in mean- ing of the words.

Roome, H. E., Towse, J. N., & Crespo-Llado, M. M. (2019). Contextual support for children's recall within working memory. *Quarterly Journal of Experimental Psychology*, *72*(6), 1364–1378. https://doi.org/10.1177/1747021818804440

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review*, *110*(1), 145–172. https://doi.org/10.1037/0033-295X.110.1.145

Salo, R., Henik, A., & Robertson, L. C. (2001). Interpreting Stroop interference: An analysis of differences between task versions. *Neuropsychology*, *15*(4), 462–471. https://doi.org/10.1037/0894-4105.15.4.462

Schnotz, W., & Kürschner, C. (2007). A reconsideration of cognitive load theory. *Educational Psychology Review*, *19*(4), 469–508. https://doi.org/10.1007/s10648-007-9053-4

Schnotz, W., & Rasch, T. (2005). Enabling, facilitating, and inhibiting effects of animations in multimedia learning: Why reduction of cognitive load can have negative results on learning. *Educational Technology Research and Development*, *53*(3), 47–58. https://doi.org/10.1007/BF02504797

Schouppe, N., De Houwer, J., Ridderinkhof, K. R., & Notebaert, W. (2012). Conflict: Run! Reduced Stroop interference with avoidance responses. *The Quarterly Journal of Experimental Psychology*, *65*(6), 1052–1058. https://doi.org/10.1080/17470218.2012.685080

Schouppe, N., Ridderinkhof, K. R., Verguts, T., & Notebaert, W. (2014). Context-specific control and context selection in conflict tasks. *Acta Psychologica*, *146*, 63–66. https://doi.org/10.1016/j.actpsy.2013.11.010

Shichel, I., & Tzelgov, J. (2018). Modulation of conflicts in the Stroop effect. *Acta Psychologica*, *189*, 93–102. https://doi.org/10.1016/j.actpsy.2017.10.007

Shimamura, A. P., Jurica, P. J., Mangels, J. A., Gershberg, F. B., & Knight, R. T. (1995). Susceptibility to memory interference effects following frontal lobe damage: Findings from tests of paired-associate learning. *Journal of Cognitive Neuroscience*, *7*(2), 144–152. https://doi.org/10.1162/jocn.1995.7.2.144

Snodgrass, Joan G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*, *6*(2), 174–215. https://doi.org/10.1037/0278-7393.6.2.174

Snodgrass, Joan Gay, & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General*, *117*(1), 34–50. https://doi.org/10.1037/0096-3445.117.1.34

Spieler, D. H., & Balota, D. A. (1996). Characteristics of associative learning in younger and older adults: Evidence from an episodic priming paradigm. *Psychology and Aging*, *11*(4), 607–620. https://doi.org/10.1037/0882-7974.11.4.607

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments & Computers*, *31*(1), 137–149. https://doi.org/10.3758/BF03207704

Stickgold, R., Whidbee, D., Schirmer, B., Patel, V., & Hobson, J. A. (2000). Visual discrimination task improvement: A multi-step process occurring during sleep. *Journal of Cognitive Neuroscience, 12*(2), 246–254. https://doi.org/10.1162/089892900562075

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*(6), 643–662. https://doi.org/10.1037/h0054651

Sutton, T. M., & Altarriba, J. (2016). Color associations to emotion and emotion-laden words: A collection of norms for stimulus construction and selection. *Behavior Research Methods*, *48*(2), 686–728. https://doi.org/10.3758/s13428-015-0598-8

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, *12*(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction*, *4*(4), 295–312. https://doi.org/10.1016/0959-4752(94)90003-5

Sweller, J. (2010). Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educational Psychology Review*, *22*(2), 123–138. https://doi.org/10.1007/s10648-010-9128-5

Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. https://doi.org/10.1007/978-1-4419-8126-4

Sweller, J., & Chandler, P. (1994). Why some material is difficult to learn. *Cognition and Instruction*, *12*(3), 185–233. https://doi.org/10.1207/s1532690xci1203_1

Sweller, J., van Merriënboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, *10*(3), 251–296. https://doi.org/10.1023/A:1022193728205

Tanaka, J. W., & Presnell, L. M. (1999). Color diagnosticity in object recognition. *Perception & Psychophysics*, *61*(6), 1140–1153. https://doi.org/10.3758/BF03207619

Towse, J. N., Cowan, N., Hitch, G. J., & Horton, N. J. (2008). The recall of information from working memory: Insights from behavioural and chronometric perspectives. *Experimental Psychology*, *55*(6), 371–383. https://doi.org/10.1027/1618-3169.55.6.371

Towse, J. N., Hitch, G. J., Horton, N., & Harvey, K. (2010). Synergies between processing and memory in children's reading span. *Developmental Science*, *13*(5), 779–789. https://doi.org/10.1111/j.1467-7687.2009.00929.x

Tse, D., Langston, R. F., Kakeyama, M., Bethus, I., Spooner, P. A., Wood, E. R., … Morris, R. G. M. (2007). Schemas and memory consolidation. *Science*, *316*(5821), 76–82. https://doi.org/10.1126/science.1135935

Tucker, M. A., Hirota, Y., Wamsley, E. J., Lau, H., Chaklader, A., & Fishbein, W. (2006). A daytime nap containing solely non-REM sleep enhances declarative but not procedural memory. *Neurobiology of Learning and Memory*, *86*(2), 241–247. https://doi.org/10.1016/j.nlm.2006.03.005

Tuckey, M. R., & Brewer, N. (2003). The influence of schemas, stimulus ambiguity, and interview schedule on eyewitness memory over time. *Journal of Experimental Psychology: Applied*, *9*(2), 101–118. https://doi.org/10.1037/1076-898X.9.2.101

Tulving, E., Schacter, D. L., & Stark, H. A. (1982). Priming effects in word-fragment completion are independent of recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *8*(4), 336–342. https://doi.org/10.1037/0278-7393.8.4.336

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review, 80*(5), 352–373. https://doi.org/10.1037/h0020071

typingMe.com. (n.d.). Retrieved December 8, 2017, from http://www.typingme.com/touch-typing-amateur/typing-tutor.php

Valdez, P., & Mehrabian, A. (1994). Effects of Color on Emotions. *Journal of Experimental Psychology: General*, *123*(4), 394–409. https://doi.org/10.1037/0096-3445.123.4.394

van Heuven, W. J. B., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). SUBTLEX-UK: A new and improved word frequency database for British English. *The Quarterly Journal of Experimental Psychology*, *67*(6), 1176–1190. https://doi.org/10.1080/17470218.2013.850521

van Merriënboer, J. J. G., & Ayres, P. (2005). Research on cognitive load theory and its design implications for e-learning. *Educational Technology Research and Development*, *53*(3), 5–13. https://doi.org/10.1007/BF02504793

van Merriënboer, J. J. G., & Sweller, J. (2005). Cognitive load theory and complex learning: Recent developments and future directions. *Educational Psychology Review*, *17*(2), 147–177. https://doi.org/10.1007/s10648-005-3951-0

van Steenbergen, H., Band, G. P. H., & Hommel, B. (2009). Reward counteracts conflict adaptation: Evidence for a role of affect in executive control. *Psychological Science, 20*(12), 1473–1477. https://doi.org/10.1111/j.1467-9280.2009.02470.x

Vergara-Martínez, M., & Swaab, T. Y. (2012). Orthographic neighborhood effects as a function of word frequency: An event-related potential study. *Psychophysiology*, *49*(9), 1277–1289. https://doi.org/10.1111/j.1469-8986.2012.01410.x.

Vernon, D. J. (2001). *Effects of colour transformations on implicit and explicit tests of memory for natural objects*. Canterbury, the UK: University of Kent.

W3Schools. (n.d.). Color names sorted by color groups. Retrieved August 25, 2019, from W3Schools website: https://www.w3schools.com/Colors/colors_groups.asp

Walker, M. P., Brakefield, T., Seidman, J., Morgan, A., Hobson, J. A., & Stickgold, R. (2003). Sleep and the time course of motor skill learning. *Learning & Memory*, *10*(4), 275–284. https://doi.org/10.1101/lm.58503

Warriner, A. B., Kuperman, V., & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods*, *45*(4), 1191–1207. https://doi.org/10.3758/s13428-012-0314-x

Wiggs, C. L., & Martin, A. (1998). Properties and mechanisms of perceptual priming. *Current Opinion in Neurobiology*, *8*(2), 227–233. https://doi.org/10.1016/S0959-4388(98)80144-X

Williams, E. J. (1949). Experimental designs balanced for the estimation of residual effects of treatments. *Australian Journal of Chemistry*, *2*(2), 149–168. https://doi.org/10.1071/CH9490149

Wixted, J. T. (2005). A theory about why we forget what we once knew. *Current Directions in Psychological Science*, *14*(1), 6–9. https://doi.org/10.1111/j.0963-7214.2005.00324.x

Wulfert, E., Roland, B. D., Hartley, J., Wang, N., & Franco, C. (2005). Heart rate arousal and excitement in gambling: Winners versus losers. *Psychology of Addictive Behaviors*, *19*(3), 311–316. https://doi.org/10.1037/0893-164X.19.3.311

Yarmey, A. D., & Paivio, A. (1965). Further evidence on the effects of word abstractness and meaningfulness in paired-associate learning. *Psychonomic Science*, *2*(1–12), 307–308. https://doi.org/10.3758/BF03343468

Yonelinas, A. P. (1994). Receiver-operating characteristics in recognition memory: Evidence for a dual-process model. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(6), 1341–1354. https://doi.org/10.1037/0278-7393.20.6.1341

Yonelinas, A. P., Aly, M., Wang, W.-C., & Koen, J. D. (2010). Recollection and familiarity: Examining controversial assumptions and new directions. *Hippocampus*, *20*(11), 1178–1194. https://doi.org/10.1002/hipo.20864

Zamorano, F., Billeke, P., Hurtado, J. M., López, V., Carrasco, X., Ossandón, T., & Aboitiz, F. (2014). Temporal constraints of behavioral inhibition: Relevance of inter-stimulus interval in a Go-Nogo task. *PLoS ONE*, *9*(1). https://doi.org/10.1371/journal.pone.0087232

Zimmerman, M. E. (2011). Speed–Accuracy Tradeoff. In J. S. Kreutzer, J. DeLuca, & B. Caplan (Eds.), *Encyclopedia of Clinical Neuropsychology* (p. 2344). https://doi.org/10.1007/978-0-387-79948-3_1247

Appendix A

List of 7 Incongruent Ink Colour Patterns and

List of Colour Names and Numbers in Each Condition (Experiment 1)

Pattern \ Word	*blue*	*brown*	*green*	*orange*	*purple*	*red*	*white*	*yellow*
1	brown	green	orange	purple	red	white	yellow	blue
2	yellow	blue	brown	green	orange	purple	red	white
3	white	yellow	blue	brown	green	orange	purple	red
4	red	white	yellow	blue	brown	green	orange	purple
5	purple	red	white	yellow	blue	brown	green	orange
6	orange	purple	red	white	yellow	blue	brown	green
7	green	orange	purple	red	white	yellow	blue	brown

Condition	Word Type	Word-Number Pairs							
Congruent	Colour	blue	brown	green	orange	purple	red	white	yellow
	Names	4	3	7	6	1	5	2	8
Incongruent	Colour	blue	brown	green	orange	purple	red	white	yellow
	Names	6	5	3	7	2	8	4	1
Neutral	Colour	blue	brown	green	orange	purple	red	white	yellow
	Names	5	1	8	4	3	2	6	7

Appendix B

List of 7 Incongruent Ink Colour Patterns and

List of 8 Colour-related Words and Numbers in Each Condition (Experiment 2 and 4)

Pattern \ Word	*sky*	*mud*	*grass*	*carrot*	*grape*	*strawberry*	*cream*	*lemon*
1	yellow	blue	brown	green	orange	purple	red	white
2	white	yellow	blue	brown	green	orange	purple	red
3	red	white	yellow	blue	brown	green	orange	purple
4	purple	red	white	yellow	blue	brown	green	orange
5	orange	purple	red	white	yellow	blue	brown	green
6	green	orange	purple	red	white	yellow	blue	brown
7	brown	green	orange	purple	red	white	yellow	blue

Condition	Pair Type	Word-Number Pairs							
Congruent	Correct	sky	mud	grass	carrot	grape	strawberry	cream	lemon
		4	1	5	7	2	6	3	8
	Incorrect (Exp. 4 only)	sky	mud	grass	carrot	grape	strawberry	cream	lemon
		6	8	2	4	7	3	1	5
Incongruent	Correct	sky	mud	grass	carrot	grape	strawberry	cream	lemon
		1	8	3	6	4	7	2	5
	Incorrect (Exp. 4 only)	sky	mud	grass	carrot	grape	strawberry	cream	lemon
		4	3	6	2	1	8	5	7
Neutral	Correct	sky	mud	grass	carrot	grape	strawberry	cream	lemon
		7	4	2	1	5	3	8	6
	Incorrect (Exp. 4 only)	sky	mud	grass	carrot	grape	strawberry	cream	lemon
		1	6	3	7	2	4	5	8

Appendix C

The List of Practice Word-Number Pairs (Experiment 3, 4, 5-B and 7)

Pair Type	Practice Word-Number Pairs			
Correct	pear	soup	song	elephant
	3	6	8	5
Incorrect	pear	soup	song	elephant
	5	3	6	8

Note. The same word-number pairs were obtained from the past studies (Hazan-Liran & Miller, 2017; P. Miller et al., 2018) for correct pairs. Incorrect pairs were different pairings of words and numbers. These pairs were shown individually and randomly in black ink.

Appendix D

List of 8 Colour Names and Numbers with Assigned Colours in Each Condition (Experiment 3)

Condition	Pair Type	Word-Number Pairs							
Congruent	Correct	blue	brown	green	orange	purple	red	white	yellow
		7	8	1	2	3	4	5	6
	Incorrect	blue	brown	green	orange	purple	red	white	yellow
		6	7	8	1	2	3	4	5
Incongruent	Correct (Ink Colour)	blue	brown	green	orange	purple	red	white	yellow
		5	6	7	8	1	2	3	4
		(brown)	(green)	(orange)	(purple)	(red)	(white)	(yellow)	(blue)
	Incorrect	blue	brown	green	orange	purple	red	white	yellow
		4	5	6	7	8	1	2	3
Neutral	Correct	blue	brown	green	orange	purple	red	white	yellow
		1	2	3	4	5	6	7	8
	Incorrect	blue	brown	green	orange	purple	red	white	yellow
		8	1	2	3	4	5	6	7

Appendix E

The List of Practice Word-Word Pairs (Experiment 5-A)

Pair Type	Practice Word-Word Pairs			
Correct	pear	soup	song	elephant
	foot	taxi	contents	rain
Incorrect	pear	soup	song	elephant
	rain	contents	foot	taxi

Note. The words on the top rows in Correct and Incorrect rows were obtained from the two original studies (Hazan-Liran & Miller, 2017; P. Miller et al., 2018). The words on the bottom rows were derived from the ANEW (Bradley & Lang, 1999). Incorrect pairs were different pairings of words and numbers. These pairs were shown individually and randomly in black ink.

Appendix F

List of 8 Colour Names and Emotion Words in Each Emotion Word Condition (Experiment 5-A)

Condition	Pair Type	Word-Word Pairs							
Positive	Correct	blue	brown	green	orange	purple	red	white	yellow
		brave	love	joyful	passion	kind	secure	happy	admired
	Incorrect	blue	brown	green	orange	purple	red	white	yellow
		secure	admired	kind	brave	passion	happy	love	joyful
Negative	Correct	blue	brown	green	orange	purple	red	white	yellow
		rage	hostile	upset	misery	cruel	unhappy	fear	lonely
	Incorrect	blue	brown	green	orange	purple	red	white	yellow
		cruel	upset	misery	fear	rage	hostile	lonely	unhappy
Neutral	Correct	blue	brown	green	orange	purple	red	white	yellow
		tighten	surge	toss	auto	depend	bleep	stretch	midair
	Incorrect	blue	brown	green	orange	purple	red	white	yellow
		auto	midair	bleep	tighten	toss	surge	depend	stretch

Note. All words were presented in black ink.

Appendix G

List of 8 Negative Emotion Words and Numbers in Each Condition (Experiment 5-B)

Condition	Pair Type	Word-Number Pairs							
Congruent	Correct	cruel	fear	hostile	lonely	misery	rage	unhappy	upset
		7	1	8	2	6	4	3	5
	Incorrect	cruel	fear	hostile	lonely	misery	rage	unhappy	upset
		4	8	5	7	3	6	1	2
Incongruent	Correct	cruel	fear	hostile	lonely	misery	rage	unhappy	upset
		6	8	5	4	7	1	2	3
	Incorrect	cruel	fear	hostile	lonely	misery	rage	unhappy	upset
		2	3	7	1	8	4	5	6
Neutral	Correct	cruel	fear	hostile	lonely	misery	rage	unhappy	upset
		1	4	6	5	3	7	8	2
	Incorrect	cruel	fear	hostile	lonely	misery	rage	unhappy	upset
		7	6	8	2	4	1	5	3

Note. All words in the congruent condition, in the incongruent condition and word-number pairs in the neutral conditions were written in white ink. The congruent ink colours were retrieved from the participants' personally selected image colours. The numbers in the incongruent condition were written in an ink colour that was randomly selected from the rest of seven unselected ink colours.

Appendix H

List of 7 Incongruent Ink Colour Patterns and

List of 8 Colour-related Words and Numbers in Each Condition on Day 1 (Experiment 6)

Pattern \ Word	*sky*	*mud*	*grass*	*carrot*	*grape*	*strawberry*	*cream*	*lemon*
1	yellow	blue	brown	green	orange	purple	red	white
2	white	yellow	blue	brown	green	orange	purple	red
3	red	white	yellow	blue	brown	green	orange	purple
4	purple	red	white	yellow	blue	brown	green	orange
5	orange	purple	red	white	yellow	blue	brown	green
6	green	orange	purple	red	white	yellow	blue	brown
7	brown	green	orange	purple	red	white	yellow	blue

Condition	Pair Type	Word-Number Pairs							
Congruent	Correct	sky 1	mud 7	grass 4	carrot 5	grape III (3)	strawberry VIII (8)	cream II (2)	lemon VI (6)
	Incorrect	sky 5	mud 4	grass 7	carrot 1	grape VIII (8)	strawberry III (3)	cream VI (6)	lemon II (2)
Incongruent	Correct	sky 6	mud 1	grass 8	carrot 3	grape V (5)	strawberry II (2)	cream VII (7)	lemon IV (4)
	Incorrect	sky 8	mud 3	grass 6	carrot 1	grape VII (7)	strawberry IV (4)	cream V (5)	lemon II (2)
Neutral	Correct	sky 2	mud 1	grass 7	carrot 4	grape VI (6)	strawberry VIII (8)	cream III (3)	lemon V (5)
	Incorrect	sky 1	mud 7	grass 4	carrot 2	grape III (3)	strawberry VI (6)	cream V (5)	lemon VIII (8)

Appendix I

List of 7 Incongruent Ink Colour Patterns and

List of 8 Colour-related Words and Numbers in Each Condition on Day 2 (Experiment 6)

Pattern \ Word	*sapphire*	*horse*	*broccoli*	*basketball*	*aubergine*	*tomato*	*baseball*	*banana*
1	brown	green	orange	purple	red	white	yellow	blue
2	yellow	blue	brown	green	orange	purple	red	white
3	white	yellow	blue	brown	green	orange	purple	red
4	red	white	yellow	blue	brown	green	orange	purple
5	purple	red	white	yellow	blue	brown	green	orange
6	orange	purple	red	white	yellow	blue	brown	green
7	green	orange	purple	red	white	yellow	blue	brown

Condition	Pair Type	Word-Number Pairs							
Congruent	Correct	sapphire	horse	broccoli	basketball	aubergine	tomato	baseball	banana
		8	2	3	6	IV (4)	I (1)	VII (7)	V (5)
	Incorrect	sapphire	horse	broccoli	basketball	aubergine	tomato	baseball	banana
		3	6	8	2	V (5)	IV (4)	I (1)	VII (7)
Incongruent	Correct	sapphire	horse	broccoli	basketball	aubergine	tomato	baseball	banana
		4	2	5	7	VI (6)	I (1)	VIII (8)	III (3)
	Incorrect	sapphire	horse	broccoli	basketball	aubergine	tomato	baseball	banana
		2	4	7	5	III (3)	VIII (8)	VI (6)	I (1)
Neutral	Correct	sapphire	horse	broccoli	basketball	aubergine	tomato	baseball	banana
		3	8	5	6	I (1)	VII (7)	IV (4)	II (2)
	Incorrect	sapphire	horse	broccoli	basketball	aubergine	tomato	baseball	banana
		6	5	8	3	VII (7)	II (2)	I (1)	IV (4)

Appendix J

List of 8 Colour-related Words and Colour Patches (in the Semantic-Priming Tasks) or Letters (in the Lexical-Priming Tasks) in Each Priming Task (Experiment 7)

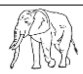
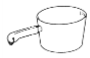
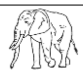
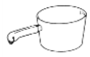
		Word-Item Pairs			
Session	Pair Type	Semantic-Priming Tasks (Ink colour name of the paired coloured patch in brackets)			
1st	Correct	aquamarine (aquamarine)	chocolate (chocolate)	orchid (orchid)	pearl (pearl)
	Incorrect	lime (bronze)	bronze (lime)	ruby (gold)	gold (ruby)
2nd	Correct	olive (olive)	amber (amber)	silver (silver)	salmon (salmon)
	Incorrect	fork (brunette)	brunette (lavender)	lavender (linen)	linen (navy)
3rd	Correct	turquoise (turquoise)	pig (pink)	marigold (marigold)	honey (honey)
	Incorrect	emerald (magenta)	amethyst (ivory)	rat (amethyst)	ivory (emerald)

		Word-Item Pairs			
Session	Pair Type	Lexical-Priming Tasks			
1st	Correct	aquamarine n	chocolate h	orchid o	pearl e
	Incorrect	lime s	bronze i	ruby z	gold b
2nd	Correct	olive l	amber b	silver i	salmon n
	Incorrect	fork a	brunette q	lavender z	linen v
3rd	Correct	turquoise s	pig g	marigold g	honey y
	Incorrect	emerald u	amethyst n	rat h	ivory f

Appendix K

The List of Practice Concept-Number Pairs

with the Sub-Groups of the Stimulus Presentation Format (Experiment 8)

Stimulus Presentation Format	Pair Type	Practice Concept-Number Pairs			
Picture	Correct	 3	 5	 7	 2
	Incorrect	 5	 2	 3	 7
Word	Correct	pear 3	elephant 5	pot 7	basket 2
	Incorrect	pear 5	elephant 2	pot 3	basket 7

Note. The pictures were retrieved from Snodgrass and Vanderwart (1980). The same word-number pairs were obtained from the studies conducted by Hazan-Liran and Miller (2017) and P. Miller et al. (2018) for correct pairs except the word *basket* and its picture. Incorrect pairs were different pairings of words and numbers. These pairs were shown individually and randomly in black ink whilst the pictures were presented monochrome.

Appendix L

List of 7 Incongruent Ink Colour Patterns (Experiment 8)

Pattern \ Word	*pig*	*horse*	*lettuce*	*carrot*	*grape*	*strawberry*	*swan*	*lemon*
1	yellow	blue	brown	green	orange	purple	red	white
2	white	yellow	blue	brown	green	orange	purple	red
3	red	white	yellow	blue	brown	green	orange	purple
4	purple	red	white	yellow	blue	brown	green	orange
5	orange	purple	red	white	yellow	blue	brown	green
6	green	orange	purple	red	white	yellow	blue	brown
7	brown	green	orange	purple	red	white	yellow	blue

Appendix M

List of 8 Colour-related Concepts and Numbers with Each Format in Each Condition (Experiment 8)

Condition	Pair Type	Picture-Number Pairs							
Congruent	Correct								
		4	1	5	7	2	6	3	8
	Incorrect								
		6	8	2	4	7	3	1	5
Incongruent	Correct								
		1	8	3	6	4	7	2	5
	Incorrect								
		4	3	6	2	1	8	5	7
Neutral	Correct								
		7	4	2	1	5	3	8	6
	Incorrect								
		1	6	3	7	2	4	5	8

Condition	Pair Type	Word-Number Pairs							
Congruent	Correct	pig	horse	lettuce	carrot	grape	strawberry	swan	lemon
		4	1	5	7	2	6	3	8
	Incorrect	pig	horse	lettuce	carrot	grape	strawberry	swan	lemon
		6	8	2	4	7	3	1	5
Incongruent	Correct	pig	horse	lettuce	carrot	grape	strawberry	swan	lemon
		1	8	3	6	4	7	2	5
	Incorrect	pig	horse	lettuce	carrot	grape	strawberry	swan	lemon
		4	3	6	2	1	8	5	7
Neutral	Correct	pig	horse	lettuce	carrot	grape	strawberry	swan	lemon
		7	4	2	1	5	3	8	6
	Incorrect	pig	horse	lettuce	carrot	grape	strawberry	swan	lemon
		1	6	3	7	2	4	5	8

Appendix N

List of Order Pattern of Face Icons in Each Session within One Colour Congruency Condition (Experiment 10)

Order Pattern	Face Icon Order		
	1st	2nd	3rd
1	☺	☹	😐
2	☹	😐	☺
3	😐	☺	☹
4	😐	☹	☺
5	☺	😐	☹
6	☹	☺	😐

Note. These six patterns repeated amongst 30 participants. 1st, 2nd and 3rd represent the number of the study-test sessions. For example, within the first colour congruency condition, the first participant saw the happy faces in the first study-test session, the second participant saw the sad faces in the first session, the third participants saw the neutral faces in the first session and so forth.