UNIVERSITY OF SOUTHAMPTON
FACULTY OF MEDICINE, HEALTH, AND LIFE SCIENCES
School of Psychology

The effects of report option and inter-target association on memory and metamemory performance in cued and uncued recall of paired associates: A generate-recognize approach

by

Mehmet Akif Guzel

Thesis for the degree of Doctor of Philosophy
September 2011

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF MEDICINE, HEALTH, & LIFE SCIENCES

SCHOOL OF PSYCHOLOGY

Doctor of Philosophy

THE EFFECTS OF REPORT OPTION AND INTER-TARGET ASSOCIATION ON MEMORY AND METAMEMORY PERFORMANCE IN CUED AND UNCUED RECALL OF PAIRED ASSOCIATES: A GENERATE-RECOGNIZE APPROACH

by

Mehmet Akif Guzel

The current research aimed to answer two main research questions. First, do variations in report-criterion option have differential effects on observing what is indeed remembered? Second, does increasing the inter-relatedness between target items have the same facilitative effect on metamemory just like it does on retrieval, or does it deteriorate in the same way as metacognitive monitoring measured by type-2 signal detection theory (dissociation)? Contrary to some earlier findings, Experiments 1 and 2 showed that participants do indeed withhold some correct responses due to the stringent report criterion. As a result, they report more correct responses when report option is maximally liberal (e.g., forced report) compared to a stringent report-criterion (free-report) particularly in uncued recall (Experiment 2). Experiment 3 found that when participants are encouraged to study cue-target pairs by focusing on targets more, inter-target association (ITA) is utilised to retrieve target items at retrieval. Thus, whilst retrieval of targets is facilitated, monitoring of the responses is not. Experiment 4 clearly showed a dissociation between memory and metacognitive monitoring due to high-ITA. Experiment 5, then, confirmed that the dissociation emerges due to the utilization of ITA by showing that it is attenuated in cued recall via ‘individuating’ the pairs (e.g., by interactive imagery). Confirming that the semantic context in which the target items studied is the critical factor to yield the observed dissociation, Experiment 6 showed that it is a strategic process that leads to the dissociation rather than solely an automatic process that facilitates retrieval of related targets by semantic activation. The results of the experiments were in line with the expectations of generate-recognize models (e.g., Bahrick, 1970) and showed that type-2 signal detection theory, which is based on this model, is an effective tool to investigate both memory and metamemory performance. The results were discussed with regards to the related literature.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

DECLARATION OF AUTHORSHIP

I, Mehmet Akif Guzel, declare that the thesis entitled, 'The effects of report-option and inter-target association on memory and metamemory performance in cued and uncued recall of paired associates: A generate-recognize approach' and the work presented in the thesis are both my own, and have been generated by me as the result of my own original research. I confirm that:

- this work was done wholly or mainly while in candidature for a research degree at this University;
- where any part of this thesis has previously been submitted for a degree or any other qualification at this University, this has been clearly stated;
- where I have consulted the published work of others, this is always clearly attributed;
- where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
- I have acknowledged all main sources of help;
- where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
- none of this work has been published before submission.

Signed :

Date : 28.12.2011

To my family

ACKNOWLEDGEMENTS

I am initially pleased to express my gratitude to my supervisor Dr Phil Higham. He has supervised my PhD work and provided me with encouragement, support, and with his invaluable ideas and suggestions. I would like to express my sincere gratefulness for his outstanding contributions to the quality of this research. I am also much grateful to my external examiner Dr Chris Moulin from the University of Leeds and my internal examiner Dr Sarah Stevenage from the University of Southampton who generously provided me with their priceless ideas and suggestions to enhance the academic quality of my PhD thesis.

I would also like to express my special gratitude to my family, Aysel Guzel, Seref Guzel, and Tahir Sahin Guzel for their endless support, sacrifice as well as their patience whilst I was away from them for a long time. Amongst my best friends, Talat Demirsoz, Olcay Sezenler, Cakil Sarac, Ahmet Coymak, Asli Tanyel, Canselin Dabanlioglu, Gizem Ayaz, Ozgun Ozturk, Berivan Aydin, Dila Karakas, Aysegul Bugra, Beste Eris, Ozlem Kaytaz, Hulya Aksoy Selen, Ahu Ozturk, Oznur Oncul, Ahmet Akif Tek, Zeynep Zelal Moroglu, Asli Kan, Sara Leuddeke-Carr, Greg J. Neil, Karlos Luna Ortega, Beatriz Martin Luengo, Davide Bruno, Elle Boag, Georgia Chronaki, and Tito Elgueta also deserve special thanks for their friendship and support throughout my years in Southampton, England.

I am also so much grateful to my funding sources, ORSAS (Overseas Research Students Awards Scheme) and the School of Psychology of the University of Southampton, for supporting me financially throughout my PhD work in the University of Southampton.

Lastly, here is my sincere thanks to the participants who rendered this research possible and provided me with interesting data.

ABBREVIATIONS USED

AMP	Accuracy memory performance
ANOVA	Analysis of variance
AUC	Area under the curve
B.A.	Backward semantic association
C-R	Cued recall
CI	Category implied (list)
CR	Correct rejection rate
F.A.	Forward semantic association
FAR	False alarm rate
Fr.A.	Free association norm
FOK	Feeling of knowing
FOL	Feeling of learning
HR	Hit rate
ISI	Inter-stimulus interval
ITA	Inter-target association
JOL	Judgments of learning
MA	Memory accuracy
MQ	Memory quantity
MR	Miss rate
NCI	No-category implied (list)
QMP	Quantity memory performance
REM	Retrieving effectively from memory model
ROC	Receiver operating characteristics
SA	Semantic association
SDT	Signal detection theory

CHAPTER 1
Literature Review

1.1. Introduction

Memory, being highly intricate subject matter in psychology, has caught the interests of a significant number of researchers since the first book describing experimental research on memory was published by Hermann Ebbinghaus in 1885 (Hunt & Ellis, 1999). After the modern cognitive era took the centre of the stage in 1960s mainly as a response to the behaviourism (Bower, 2000), the number of studies on cognition increased in number and scope. Amongst those areas, a new investigation topic appeared in cognitive psychology in recent years and has taken the interests of many scholars, which is metacognition.

The six experimental studies reported throughout this thesis in detail are particularly concerned with shedding more light on memory performance together with metacognitive processes in cued and uncued recall of paired-associates. The backbone of the research lied on the basis of generate-recognize perspective (e.g., Bahrick, 1968, 1970; Anderson & Bower, 1972), and so the current research fundamentally aimed to investigate the memory along with metamemory performance in recall via considering the assumptions of this theory.

More specifically, the present research aimed to answer two main research questions. First, does varying the report option that participants adopt at the time of testing affect what is truly remembered? In other words, do people report fewer correct responses under lenient report criterion (e.g., free report) compared to when they adopt stringent report option (e.g., forced report)? This question is important to test one of the fundamental assumptions of the generate-recognize approach (Bahrick, 1969, 1970; Kintsch, 1970). Proving that the basic assumption of the approach is valid, participants should be expected to generate not only correct candidates but also incorrect ones so that greater number of correct responses should be gathered when a stringent report criterion is employed compared to when a liberal report option is adopted. In relation to this question, the research question also aimed to understand which possible factors might lead to the observation that participants report more correct responses when report option is lenient than when it is stringent. Second, does organization in memory (e.g., categorization) have the same facilitative effect on metacognitive performance such as monitoring as it has on memory performance? This question is important mainly because the literature confirms that organization in memory facilitates retrieval performance, however, the effects of organization on metacognitive processes (e.g., on metacognitive monitoring) are unclear. Hence, both of the questions above have a common investigation. Both of the questions intended to investigate whether generate-recognize approach could be taken as a valid approach in measuring not only memory performance but also metacognitive performance, although the model has been criticized much and was hindered to develop not later than it was proposed, particularly by the encoding-specificity principle (Thomson & Tulving, 1970).

This chapter, therefore, starts with defining metacognition and the metacognitive processes and proceeds to lay out some of the recent available approaches to measure metacognitive processes; the strategic regulation of memory accuracy framework of Koriat and Goldsmith (1996c) and type-2 signal detection theory (i.e., Banks, 1970; see also Higham, 2002). Owing to the specific subject matter of the present study and in better relation to the research questions, the chapter will continue to review one of the theories of retrieval in detail, which is generate-recognize theory (e.g., Bahrick, 1969, 1970) along with the criticisms directed to it, particularly by the recognition failure of the recallable words phenomenon (Thomson & Tulving, 1970). Lastly, the chapter will elaborate on the rationale and the aims of the present research by addressing the above-mentioned research questions in terms of the related literature. The overview of the experiments will be given at the end of the chapter.

1.2. Metacognition and metacognitive processes

In recent investigations on memory, metacognition has become a popular area and it has underlined the opinion that memory should not be considered as a simple subject matter that involves some mechanical structures and simple serial processes. Rather, it has some more complex processes, such as metacognitive processes guiding cognitive behaviour and affecting memory performance. In this sense, *metacognition* is conventionally defined as the knowledge and experiences we have about our own cognitive processes (Flavell, 1979).

The modern understanding of metacognition appeared in cognitive psychology after the publication of Nelson and Narens (1994) paper on two key metacognitive processes: metacognitive monitoring and control. *Metacognitive monitoring* is referred to as those processes allowing individuals to observe, reflect on, or experience their own cognitive processes. For instance, one may feel that he or she has understood the text just read, or has a feeling-of-knowing (FOK), feeling-of-learning (FOL), or judgements of comprehension that are observed in laboratory settings. *Metacognitive control*, on the other hand, refers to the conscious and unconscious decisions that we make based on the output emerged as a result of our monitoring processes (Perfect & Schwartz, 2002). As a relationship between these two metacognitive processes, metacognitive control is the ability to use the judgements made by individuals about their own cognitive processes (monitoring) so as to alter the behaviour (Nelson & Narens, 1994).

The interdependency between metacognitive monitoring and control processes was developed into a theory by Nelson and Narens (1994); see Figure 1. The theory involves two structures: a *meta-level*, operating as a dynamic process in which it works by the assessment of the current state and is guided by the introspection, and an *object-level*, including the actions and behaviours of the individual. According to Nelson and Narens, meta-level is informed by the object-level during metacognitive monitoring, and meta-level modifies the object-level during metacognitive control. Consider a memory test where participants are free to report the words they remember from a just presented list. That is, participants may choose to report or withhold giving a response. When a participant tries to recall a target item from the presented list and if he/she has a vague idea that the remembered word is a word presented in the list or thinks that it is indeed not a studied item (good monitoring), the participant will most probably withhold that item (control). Based on the theory of Nelson and Narens, the metalevel is informed by the object level that the word has not been remembered (or vaguely remembered), which refers to monitoring process. As a control process, the person withholds giving a response.

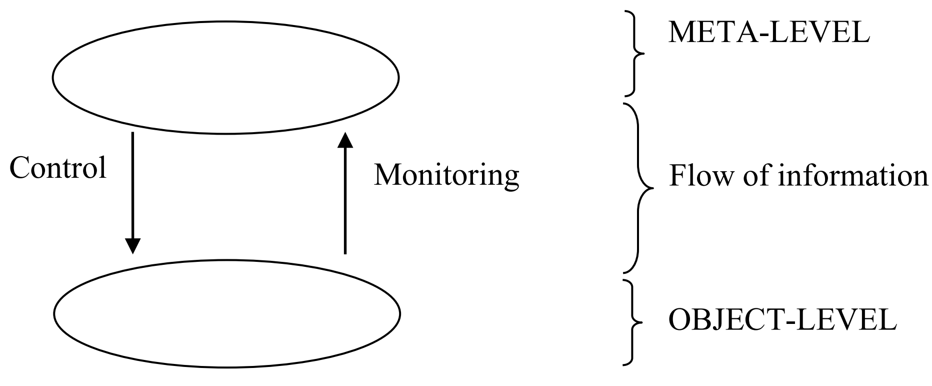


Figure 1. Nelson and Narens' (1994) schematic formulation of two-level structure for metacognitive mechanisms (Source: Nelson, O. Thomas and Narens, L. (1994). Why investigate metacognition? In J. Metcalfe, and A. P. Shimamura (Eds.), Metacognition: Knowing about knowing, p.1-25. Cambridge, MA: MIT Press).

Metacognitive processes have been investigated in various settings. For instance, in a developmental setting, Roeder (2006) investigated the ability of children to strategically regulate their memory accuracies and showed that competencies on strategic regulation (despite being slow) develops continuously during primary school years and shows the emerging signs from age of seven. In an educational domain, Hacker, Bol, Horgan, and Rakow (2000), for instance, asked students to predict the percent of correct responses they would get on a multiple-choice exam. The correlation between the predictions made and the performance of the students yielded that it increased with test experience for those who performed well on the exam. Another popular application of metacognition in the cognition literature is eyewitness memories (e.g., see Roberts & Higham, 2002; Granhag, Jonnson, & Allwood, 2004; Allwood, Granhag, & Jonnson, 2005, 2006) in which the importance of cognitive interviews and their practical implications on eyewitness memories are investigated.

Whilst some research has observed monitoring prospectively where the monitoring performance of the participants are gathered before the testing is conducted, such as by the judgements-of-feeling or judgements-of-knowing (e.g., Hart, 1965, 1966), or feeling-of-learning (e.g., Nelson & Leonesio, 1988), some others have measured monitoring retrospectively such as by confidence judgements given to responses (e.g., Koriat & Goldsmith, 1996c; Higham, 2002; Sauer, Brewer, & Weber, 2008).

1.3. Measuring metacognitive processes

In better relation to the current research, the following sections will review two of the currently available methods to measure monitoring performances: the framework of Koriat and Goldsmith (1996c) and type-2 signal detection approach, in both of which the measurement of monitoring performance mainly rests on the retrospective judgements on responses given (e.g., confidence levels).

1.3.1. *Strategic regulation of memory accuracy: Telling less for the sake of being more accurate in what you tell*

People are proposed to have an ability to strategically regulate their memory accuracy performance. In regard to such metacognitive regulation, the framework of Koriat and Goldmith (1996c) seems remarkable and their framework is practically valuable as a means to investigating memory performance along with metacognitive processes. The framework was proposed to investigate controlled-experimental studies in its original form, however, Koriat and Goldsmith essentially adapted the ideas for strategic regulation of memory accuracy from everyday memory such as the swearing in traditional court cases that involves a report criterion for the witnessed event: "to tell the 'whole' truth, but nothing but the truth" (Goldsmith & Koriat, 2008, p. 46).

Koriat and Golsmith (1996c) proposed two different measurement methods in order to observe a better as well as a clearer measurement of memory performance in laboratory settings: *quantity memory performance* (QMP) and *accuracy memory performance* (AMP). The quantity memory performance refers to the proportion (or percentage) of the number of correctly recalled items out of total number of items 'studied'. On the other hand, accuracy memory performance refers to the proportion (or percentage) of the correct responses out of total number of items 'reported' (Koriat & Golsmith, 1994, 1996c). For instance, if a participant studies 20 words in a list-learning experiment and recalls 12 words amongst which 10 are correct, input-bound QMP of this participant is .50 (10/20). That is, 50% of the material is successfully recalled. This performance was termed as *input-bound memory performance*, because the performance is conditional upon the amount of presented information at the time of study that is retained and is accessible at the time of test (input). On the other hand, when this participant reports 12 words amongst which 10 are correct and two are incorrect, the AMP of this participant is .83 (10 /12). In other words, 83% of the answers are, in fact, correct. Therefore, whereas QMP was best described with a *storehouse metaphor* since memory was seen as a storehouse in which discrete items of information are deposited first and then retrieved in later occasions, AMP refers to the *output-bound memory performance* and bases on their *correspondence metaphor*. As a result, QMP and AMP together complement the measurement of strategic regulation of memory accuracy and informativeness (Koriat & Goldsmith, 1996c; see also, Goldsmith & Koriat, 2008). According to Koriat and Goldsmith, the distinction between QMP and AMP is important because the AMP itself reflects the dependability of the information reported. This dependability was the degree to which each item reported by the participant can be trusted to be correct. Hence, whilst the rememberer himself/herself is being held responsible for what he or she fails to report, the output-bound accuracy measure holds the person liable only for those he or she reports (Koriat & Goldsmith, 1996c).

Further, QMP and AMP measurements can be measured and informative only for the experimental designs that use free-report option (Koriat & Goldsmith, 1996c). *Free-report* refers to the option that is given to the participants and allows them to be free in withholding any responses at the time of testing. On the other hand, *forced-report* refers to the conditions in which participants are asked to give a response (to every item or question) even if they need to make guesses. Hence, input-bound QMP and output-bound AMP are necessarily equivalent when a study utilises only forced-report option. This is mainly because the numbers of studied and reported items are equal as the participant is forced to report all of the studied items.[1] Therefore, these measurements remain as a matter of interpretation as the score gathered reflects either the input-bound or output-bound performance in forced report. However, the number of output items (number of totally reported items) might be fewer than the number of input items (number of totally studied items) in the studies having free-report option where the participants are provided with a choice to say; 'I do not know/remember'. Hence, unless the participant reports all of the words correctly or if monitoring is at chance, it is much more probable for the participants to have higher AMP than QMP free-report memory experiments (Koriat & Goldsmith, 1996c).

According to Koriat and Goldsmith (1996c), participants are capable of strategically regulating their memory performance for the sake of being more accurate in their responses. That is, when participants are given the option to withhold giving an answer such as under free-report options, they strategically refrain from giving the responses that they consider incorrect. Therefore, this strategic regulation mainly reflects itself in yielding higher AMP than QMP. Figure 2 presents the model of Koriat and Goldsmith (1996c), which explains how the metacognitive processes operate in regulating AMP and QMP through the following processes in order: retrieval, monitoring, and control.

Although the framework of Koriat and Golsmith (1996c) borrowed much from the signal detection theory (SDT), they criticized the signal detection approach for not being capable of measuring metacognitive processes during retrieval (Koriat & Goldsmith, 1996c). However, signal detection theory was also proposed by some researchers (e.g., Higham, 2002; see also, Higham & Tam, 2005, 2006) as being a fruitful method to measure metacognitive processes (e.g., see Higham, 2011, for a further review and discussion). Therefore, the application of the theory on measuring metacognitive processes will be laid out in detail in the following section.

[1] For instance, if a participant is forced to report 20 words after studying a list of 20 words and 15 of them are correct and five are incorrect, QMP and AMP are 'both' 75%.

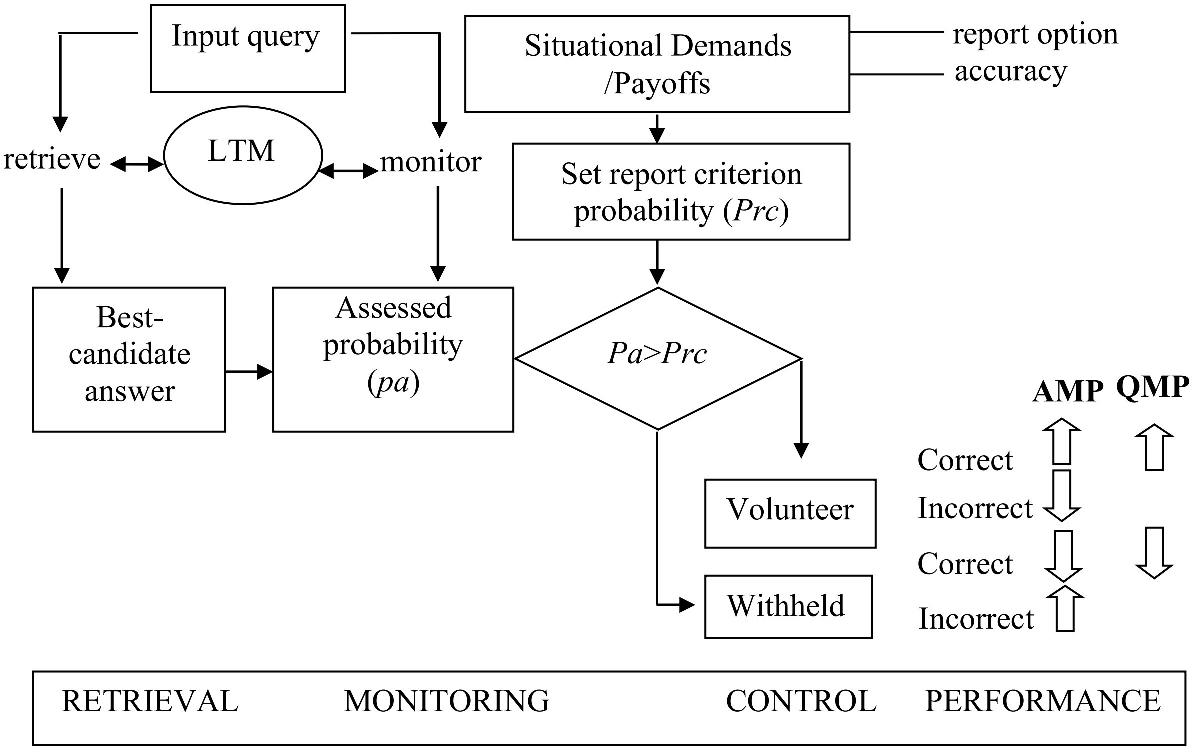


Figure 2. Koriat and Goldsmith's (1996c) model of strategic regulation of memory accuracy (Source: Koriat, A. and Goldsmits, M. (1996c). Monitoring and Control Processes in the Strategic Regulation of Memory Accuracy. Psychological Review, 103 (3), 490-517.)

1.3.2. *Application of signal detection theory on measuring metacognitive processes*

Signal detection theory was originated by Green and Swets (1966) in their book, "Signal Detection Theory and Psychophysics". The theory initially entered into the literature of psychology as a method to explain the decisions on the detection of sensory stimuli that, for instance, require distinguishing weak visual or auditory signals from a noisy background. Therefore, it provides a general framework to describe and to investigate the decisions that observers make in ambiguous, uncertain situations (Macmillan & Creelman, 2005). Although it is widely-used in *applied psychophysics*, which is an area studying the relationships between a physical stimulus and its psychological (or subjective) effects, it has also been considered as a theory which have quite informative implications on how any type of a decision is made in uncertain situations (Wickens, 2002). For instance, it was utilised in yes-no recognition tasks in which participants are asked to make decisions in a mixed list of old (studied) and new (unstudied) items as to whether the items have been studied or not (Abdi, 2007).

Based on the theory, there are four possible rates that could be calculated: *hit rate* (*HR*), *false alarm rate* (FAR), *miss rate* (MR), and *correct rejection rate* (CR); see Table 1. For instance, in a word recognition test in which the participants are asked to say 'yes' or 'no' for the occurrence of each of the words at study, a contingency table is drawn between participants' response (yes-no) and the reality (old-new), then the above-mentioned rates are calculated accordingly. As it is illustrated in Table 1, HR of a participant is the rate of correctly detecting the words when she/he has actually been presented with at the time of study (e.g., participant says 'yes' for the 'old' items). FAR, however, is the rate of responding 'yes' to the words that were actually absent in the studied list. CR, as the term implies, occurs when the participant says 'no' when the word was absent at study. Lastly, MR is the rate of responding 'no' to the words actually presented (old items).

Table 1

Four Possible Rates that can Occur in a Yes/No Recognition Test

	Participant's response	
Reality	"Yes"	"No"
Old (signal present)	Hit (%)	Miss (%)
New (signal absent)	False alarm (%)	Correct rejection (%)

Source: Klatzky, R. L. (1975). Human Memory: Structures and Processes, p. 245, W.H. Freemand and Company, San Francisco: USA.

The goal of the theory is to measure the parameters such as d-prime (d') and beta (β) from the experimental data; see Figure 3. D-prime (d') indicates the relative strength of the signal to the noise assumingly distributed along two normal distributions and it is measured as the distance between the means of these two normal distributions in standard deviation units. Therefore, it is inferred that the more participant learns the items (in other words, makes an effective discrimination between old and new items), the further apart are the means of the two distributions. The criterion called beta (β) indicates the participants' response strategy, and it is the strength criterion of the participant on which she/he bases the decision (Klatzky, 1975).

More importantly, however, signal detection theory has been proposed to have two types: *type-1* and *type-2*. The clarification on this distinction is important to understand the prospective application of type-2 SDT on metacognitive research.

Whereas type-1 analysis is defined as the one applied when participants evaluate external stimuli provided by the experimenter, type-2 SDT has been applied when participants evaluate their own responses (Clarke, Birdsall, & Tanner, 1959). Therefore, participants using type-1 SDT decisions might evaluate the external stimuli provided to them with such as a binary decision as to whether the items were present in the studied list. The participant may use the binary decision procedure via responding as ‘signal /noise’ in a signal detection test, or ‘old/new’ in a recognition test, or he/she may use the confidence rating procedure by which he/she indicates his/her confidence that the signal was present, or the item in the recognition test was an old item. This type of SDT application has been utilised overwhelmingly by the research that concerns the decision processes on such as visual or auditory sensory stimuli (e.g., cf. Antrobus & Singer, 1964; Miller & Leibowitz, 1976; Chiarello, Liu, Quan, & Shears, 2000; Sanabria, Spence, & Soto-Faraco, 2008).

In type-2 SDT approach, however, participants evaluate ‘their own responses’ that they have just given at the time of testing; see Figure 4. In type-2 SDT analysis, the evaluations can be made in terms of a binary decision as ‘correct/error’ as well as in terms of the levels of confidence that indicate *how participants judge the correctness of their own responses*. Even though the analysis in binary decisions is a judgement having two values, the evaluation in confidence ratings can have multiple values of a stimulus or the responses given (Clarke, Birdsall, & Tanner, 1959; see also Banks, 1970).

To sum up, whilst the distributions base on the actual correct and incorrect responses in type-1 SDT approach, the distributions are of the responses that were judged to be correct and those judged to be incorrect. Therefore, the observer decides which of the two (or more) events has occurred that are defined independently of the observer himself or herself. However, another event occurs at the time of type-1 decision: The observer is either correct or incorrect in his or her decision. Hence, the task of the observer discriminating between his/her correct and incorrect type-1 decisions is called a type-2 task (Clarke, Birdsall, & Tanner, 1959). As a result, since it is response-contingent, type-2 SDT modelling provides more information on the processes that are made at the time of retrieval (e.g., the subjective confidence levels given on the correctness of the responses given) unlike stimulus contingent type-1 SDT decisions.

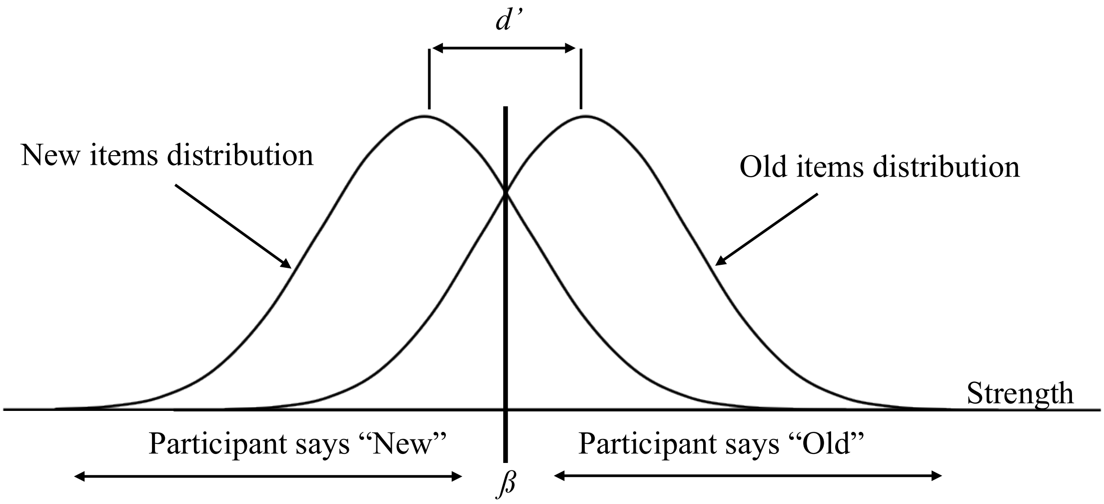


Figure 3. Distributions of old and new items along the continuum of strength assumed to happen in the internal state of participants on the basis of type-1 (stimulus-contingent) SDT modelling. The vertical line represents the criterion (*ß*), and *d'* indicates discrimination index.

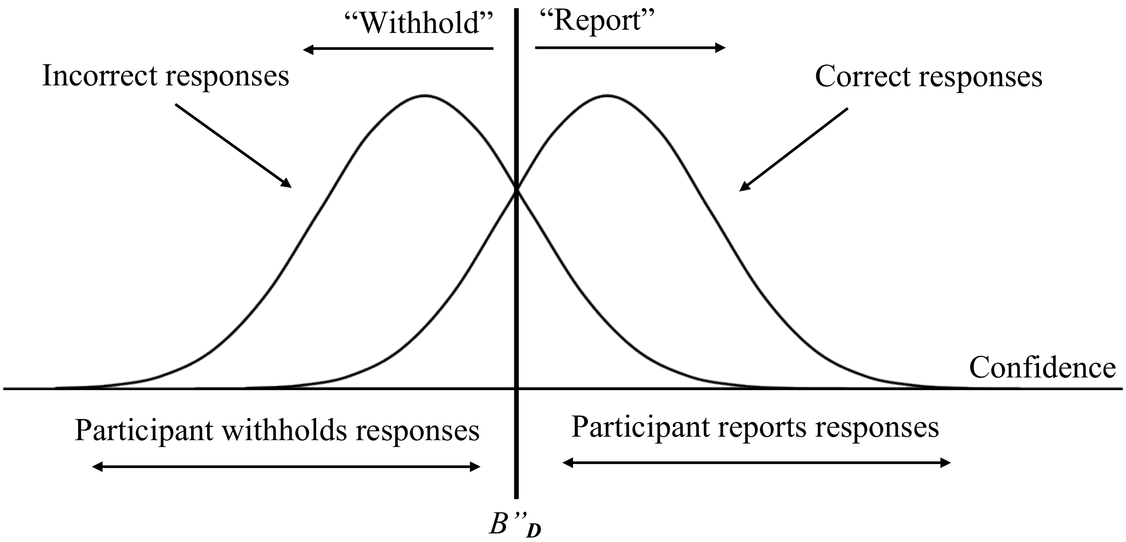


Figure 4. Distributions of correct and incorrect responses along the continuum of confidence assumed to happen in the internal state of participants on the basis of type-2 (response-contingent) SDT modelling. The vertical line represents the report criterion (*B"D*) above which candidates are reported below which candidates are withheld. (Source: Higham, P. A. & Arnold, M. M. (2007). How many questions should I answer? Using bias profiles to estimate optimal bias and maximum score on formula-scored tests. European Journal of Cognitive Psychology, 19 (4/5), 718-742).

The common idea employed by the framework of Koriat and Goldsmith (1996c) as well as type-2 signal detection theory lies on the fact that both of the approaches consider the basic assumptions of generate-recognize theory in the retrieval stage. For instance, the input query of stage of the Koriat and Goldsmith's strategic regulation of memory accuracy framework (1996c) involves a generation process of best candidate answer(s) which is followed by a process of monitoring the correctness of the answer generated (retrieved), which is termed as assessed probability; see Figure 2. Parallel to this approach, for instance, Higham (2002; see also Higham & Tam, 2005, 2006) proposed that the type-2 signal detection theory approach could be well suited to measure metacognitive processes (e.g., metacognitive control, response bias, and monitoring). As was suggested by Higham and Tam (2005), the scores of *monitoring* (e.g., *A'*) and *report bias* (e.g., *B"D*) are calculated on the basis of a contingency table which is drawn to calculate the frequencies of the numbers of correct and incorrect responses that are reported and are withheld (e.g., see section 2.3.4. and Table 4). *A'*, as a measure of monitoring, was defined as the degree to which participants have a tendency to report correct candidate responses and withhold incorrect ones. Report bias, however, was defined as the tendency of participants' to report the candidate responses (or to say 'yes' in 'old/new' recognition task) regardless of their accuracy (see e.g., Higham, 2002). Therefore, the application of type-2 signal detection theory makes a likening between 'generation' followed by a recognition process and 'recalling' (or retrieving) the stored information from memory followed by a monitoring performance as to whether the remembered (generated) candidates are correct or not.

The following section will review the generate-recognize theory in detail, both because generate-recognize theory plays a critical role in understanding the applications of some of the currently available measurement strategies of the metacognitive processes mentioned above and the research questions are based on testing the assumptions of theory.

1.4. Generate-recognize theory of recall: recognizing the correct response from amongst the generated candidates

The *generate-recognize theory of recall* (two-stage or dual-process theory) has its essence in making a distinction between generation and recognition processes whilst retrieving information from memory. The theory has its forerunners as Kintsch (e.g., 1968), Bahrick (1970), and Anderson and Bower (1972) who developed it in the context of list learning. The theory basically proposes a permanent knowledge system in which each atom (or unit) corresponds to a different idea or concept and some of them, but not all, are labelled as 'words' (Watkins & Gardiner, 1979). In a specific proposal, encountering a word (i.e., presenting words at the time of study in a recall experiment) provides an automatic access to the representation of the word in the cognitive system which results in an attachment of an 'occurrence tag'. On a later recall test, however, the theory suggests that a 'search mechanism' is made throughout the cognitive system, and then, some possible candidate representations of words are 'generated'. Following this stage, each generated representation of the word is subjected to a process of 'recognition' during which a decision as to whether an occurrence tag exists or not. Following the suggestion of the theory, if the test is recognition rather than a recall test, an encounter with the test word guarantees the access to the representation of the word. Therefore, a failure in recall is attributed to either the representation of the word has not been generated or a wrong (recognition) decision process has been made. However, if a failure happens in recognition, the reason of that failure is attributed only to the wrong recognition decision made (Watkins & Gardiner, 1979).

The proposal of the theory can be illustrated with the following statement of Bahrick (1970):

> For example, the individual who wants to recall the name of the girl he took to a high school prom, but fails in his initial effort, may produce a list of girl's names from his general memory store until he retrieves a name which he recognizes as the correct one. If he also happens to remember that the name began or ended with a certain syllable and restricts the repertoire of retrieved names in accordance with these cues, the likelihood of successful retrieval is increased. (p. 215)

The above-mentioned generate-recognize route was taken as a counter alternative to *direct access* view, which proposes that when information is attempted to be retrieved from memory it simply results in either accessing the stored information or a failure in doing so. However, Bahrick (1969, 1970) proposed that the generate-recognize route was only appropriate to be implemented if direct retrieval is failed since he makes a compromise that there may not only be a single route as generate-recognize proposal assumes but there might also be a direct route. Bahrick (1970) showed empirical evidence for the two-stage retrieval process via gathering a greater memory performance when prompting (providing cues) participants for the items that they failed to recall in the first trial. Based on this finding, he proposed that those directly recalled in the first attempt, where he thought that the direct route was utilised at the first instance, could be recovered afterwards by generate-recognize route to access the sough-after information (Bahrick, 1970). Hence, these two possible routes of memory access, direct-access versus generate-recognize route, in terms of conscious recollection might be seen as complementary rather than contradictory to each other.

Generate-recognize theory has also undergone some modifications (e.g., see Higham, 2002 for a review). In its early traditional form, cued recall was proposed to be a process which could be achieved by covertly generating some ‘associates of cues’ available at the time of test, and then trying to recognize the target from amongst the generated candidates (Kintsch, 1968; Bahrick, 1970; Anderson & Bower, 1972). The theory had some empirical support from the studies of conducted by researchers, such Bahrick (1970) who demonstrated the probability of recall was affected by the associative strength between cue and target. However, Jacoby and Hollingshead (1990) offered a pivotal change to the early versions of generate-recognize model. They suggested that assuming the source of candidates in which they are generated is an abstract and stable network (cf. Bower, 1980) could be changed with the assumption that the source is distributed rather than a stable associative network. For instance, Jacoby and Hollingshead considered that generating candidates in order to recognize the sough-after item from amongst the generated candidates are influenced by specific episodes. Therefore, the source of generation was mainly assumed to be based upon episodic memory rather than semantic memory (Jacoby & Hollingshead, 1990).

The theory has attracted its critics as well and it was hindered to develop before it became mature enough to tackle with various memory queries. The most prominent criticism directed to generate-recognize theory came from the experiments of Endel Tulving and his colleagues (i.e., Tulving & Olser, 1968; Thomson & Tulving, 1970). Before the research questions of the current research are reached, the following section will review the scientific game between encoding specificity hypothesis and generate-recognize theory of recall.

1.5. A scientific battle: encoding specificity against generate-recognize theory

As a critic against generate-recognize theory, Thomson and Tulving (1970) aimed to test the *associative continuity hypothesis* (Bilodeau & Blick, 1965). According to this hypothesis, the effectiveness of cues was attributed to the extent of the semantic association between the cue and the target (e.g., strong vs. weak). It was proposed that the strong semantic association between cue and target pair (e.g., table–chair) would turn out to be an effective aid in retrieving the target (chair) paired with it 'even if' the cue would be absent at study (Bilodeau & Blick, 1965).

However, Thomson and Tulving (1970) found contradictory results that should not be expected based on the associative continuity hypothesis, which lay on the assumptions of generate-recognize theory. Thomson and Tulving (1970) demonstrated the following. When the target items studied with their weak-associate cues (e.g., shoes-CHAIR) and the target items are asked to be recalled in the context of the cues studied (shoes-?), strongly associated extra-list cues provided at the time of test (i.e., table-?) were not found as effective as weakly associated yet studied with the target word together (shoes) (Thomson & Tulving, 1970). According to the generate-recognize theory, it should be predicted that the strong associates of target would be superior cues for recall since the probability of generating the correct target would be high.

The support for the criticism of Thomson and Tulving (1970; see also, 1973) against generate-recognize approach was termed in the cognitive psychology literature as the *recognition failure of recallable items phenomenon*. Thomson and Tulving (1970) demonstrated this phenomenon with a two-phased experimental design. In the first phase, participants were asked to free associate to the extra-list strong associates of the targets after studying weakly associated cue target pairs (generation). Then, they were asked to indicate (e.g., circle) the target words amongst those covertly generated words (recognition), which would involve the targets with a high degree of probability. This was the generate-recognize phase of the experiment. In the second phase, participants were asked to recall the target items with the help of weak-associate cues that had been studied previously. The results revealed that the targets that had not been recognized during generate-recognize phase were later recalled with the help of weak cues studied with the targets. After several replications of the recognition failure of the recallable words phenomenon (e.g., Thomson & Tulving, 1973; Tulving, 1974), the phenomenon laid the Wiseman-Tulving Law (cf. Wiseman & Tulving, 1976). Failing to find particular results which should have been expected on the basis of generate-recognize view of recall, Tulving and his colleagues (e.g., Tulving & Thomson, 1970; Thomson & Tulving, 1973, 1974; Wiseman & Tulving, 1976), hence, proposed an alternative explanation for the phenomenon: encoding specificity principle (e.g., Tulving & Thomson, 1971).

The encoding specificity principle designed in the above-mentioned phenomenon basically referred to the observations where the retrieval performance of an item is higher when the cue(s) of the to-be-remembered available at study is also available at testing compared to the cases where there exists no overt cues or not the same cues at testing (see e.g., Thomson & Tulving, 1970), has been inspected in various experimental studies as well (e.g., Bartling & Thomson, 1977; Newman, Cooper, Parker, Sidden, Gonder et al., 1982). For instance, Newman et al. showed that when cues (either strong or weak associates of target items) presented at encoding were also presented at retrieval (cued recall) were found to be more effective in recall than the recall performance of uncued participants. It should, however, be noted that the extra-list strong cues were found more effective than having no cues at test regardless of the amount of pre-training, type of instructions, length of time available for encoding, number of encoding trials, and length of interval between encoding and retrieval (Newman et al., 1982).

1.5.1. *The critics against encoding specificity hypothesis*

Despite the fact that it had strong experimental support, proposal of encoding specificity principle particularly as a criticism for generate-recognize model was also contested. For instance, it was thought that assuming *trans-situational identity of words* (Thomson & Tulving, 1973), which refers to the assumption that a single representation exists for each word in the cognitive system, should not be taken as a criticism for generate-recognize theory (e.g., Martin, 1975). That is, the meanings of the generated items at the stage of generated-recognize phase of the experiments of Tulving and his colleagues (e.g., Thomson & Tulving, 1970, 1973) could be different than the meaning of the targets studied. To be more specific, as Martin reasoned, the targets could be interpreted with another sense(s) of them that were not the same as the sense of the targets primed with weak cues at study. Therefore, recognition of the targets, which are generated at the generation phase of the recognition failure of the recallable words phenomenon and are 'nominally' the same target items studied previously, would naturally be difficult (Martin, 1975). For instance, the word 'light' cued with a weak-cue word 'head' may not be recognized when it is generated with a strong cue 'dark'. In other words, whereas the studied one (e.g., head-LIGHT) has the meaning of an article of 'furniture', the generated one has the meaning of 'luminance' (e.g., dark-LIGHT) (Martin, 1975). Therefore, being able to recognize the target (e.g., 'light') should not be expected, although the generated target and the one to be recognized are nominally identical.

In the same vein, Reder, Anderson, and Brojk (1974) also considered the alternative of multiple representations of words rather than single-representation. Reder and his colleagues suggested that the interpretation of the generated words for extra-list strong cues were different than the words paired with weak cues at study. In order to test this suggestion, Reder et al. manipulated the frequency of the words utilised at study (based on the normative data on written frequency of words) and they found that encoding specificity was valid for only the words used frequently. However, the encoding specificity hypothesis, which was tested in recognition failure paradigm, was not supported for the words used less frequently (Reder et al., 1974). Their results specifically showed that although generation possibility of frequent words was higher than recognizing these words, infrequent words were generated (recalled) equally well as they were recognized. This difference was, then, attributed to the fact that the infrequent words have relatively ‘less (number) senses’ compared to the frequent words, and so being able to recognize the target from amongst the generated candidates would be easier for the infrequent ones. Following Reder, Anderson, and Bjork, although Tulving and Watkins (1977) found recognition failure effect with words having single meaning, Muter (1984) studied the recognition failure of recallable words with famous and unique names. However, Muter could not find recognition failure for the unique names.

Additionally, some researchers stressed the importance of report option (e.g., forced vs. free-report options) in the experiments concerning the recognition failure. For instance, Pellegrino and Salzberg (1975; see also, 1974) used yes-no type of recognition tasks along with a forced-report option. They supported the encoding specificity principle by showing that the effectiveness of functional cue-target match requires the availability of cues both at the time of input (study) and output (test). However, they conceptualized the results with the notion of *feature sampling*. Although they favour the generate-recognize model, the *feature sampling theory* (e.g., Pellegrino & Salzberg, 1975) can be regarded as another conceptualization of the encoding specificity: Sampling the features of the to-be-remembered (t-b-r) items at the time of study have an improvement in retrieval (in their case, recognition performance) when these features were available at input as well.

Further, Santa and Lamwers (1976) criticized the results of the experiments conducted by Tulving and his colleagues (e.g., Tulving & Osler, 1968; Tulving &

Thomson, 1970) with the following reasons. They proposed that the recognition failure could be attributed to the switch between the instructions in those experiments that could have created confusions for the participants. That is, in typical recognition failure experiments (e.g., Thomson & Tulving, 1970), the participants were instructed at study that they would be asked to remember the target words with the help of studied cues. However, they were instructed differently at test in a way that they were asked to generate items any related items to the (extra-list) strong cues given, and then they were asked to recognize which of the generated ones were studied. Hence, Santa and Lamwers (1976; see also 1974) proposed that the improvement could be achieved with strong extra-list cues via 'informing' the participants that these (extra-list) cues are strong associates of the targets. Besides, Santa and Lamwers (1976) suggested that the participants in recognition failure experiments are asked to recognize the words amongst these generated candidates, however, the items generated by the participants themselves were 'closely related to each other' to the greatest extent. Therefore, according to Santa and Lamwers (1976), that is why recognizing the target(s) amongst these close-associate candidates turns out to be a difficult task for the participants. However, against this confusion proposal, Wiseman and Tulving (1976) showed that the recognition failure occurred even the recognition sub-phase of the experiment is 'experimentally provided', in which the generated items were designated as weakly associated to each other, rather than they were 'self-generated'.

Salzberg and Pellegrino (1974) investigated the recognition failure of recallable words phenomenon from the perspective of signal detection theory (SDT; see section 1.3.2. for the details of the theory). They investigated the false alarm rates and proposed that the generated items are not recognized but later recalled with specifically-encoded weak cues seems to be attributed to the *response bias*. As it is implied by the term, response bias refers to the tendency to respond on some basis rather than other as a form of favouritism such as the tendency to say 'yes' in a yes-no recognition task (Macmillan & Creelman, 2004). For instance, false alarm (FAR) rate for different category words, referring to the category switch of targets studied and tested, was 22.5%; However, FAR for the same category sets of words were 9.5% (Salzberg & Pellegrino, 1974). Schwartz (1975) also detected the response bias in recognition failure experiments. He designed his experiments in a way that the participants were allowed to study the target material with weak associates and then they were asked to generate and recognize the targets with extra-list strong associates, however the targets were manipulated as being either common or rare words. Recognition failure was only found for common words, but not for rare ones. It was found that although retention was same (measured with *d'*), response bias was different: Common cues resulted in higher response bias (Schwartz, 1975).

In some other experiments, providing extra-list strong cues was found to be more effective than the uncued-recall conditions. For instance, Baker and Santa (1977) compared the free-recall (uncued-recall) performance and the cued-recall performance with extra-list cues. They found higher memory performance for the strong extra-list cues than the uncued-recall conditions. The reason for such a difference seemed to depend on the elaboration level that the participants engaged at study. For this allowance, Baker and Santa had participants studied the target words with weak cues either in a shallow processing (simple pairing) or in a deep processing (targets embedded in sentences). They found that the greater the integrated context, the less the degree of benefit of extra-list cues. As a result, Baker and Santa reasoned that when an encoded context is well integrated, non-encoded retrieval cues (i.e., extra-list strong cues) are relatively less effective. Their investigation was important to find a contradiction to the encoding specificity principle by showing that extra-list strong cues could be more effective than having no cues. Their findings were also important to provide an implication for learners. That is, when students integrate the material too much to specific contexts, they have less ability to utilise other sources as contextual cues to gain an access for the sought-after material (Baker & Santa, 1977).

In regard to some other challenges against encoding specificity in the literature, beside finding such as higher extra-list cue effectiveness over uncued recall (e.g., Baker & Santa, 1977), Lauer (1974), for instance, found that input organization improved the memory performance of uncued-recall performance. Lauer varied the study-list material in terms of either having blocked category exemplars, or blocked category exemplars in alphabetic orders, or random words. It was found that clustering the material had an improvement on uncued-recall performance and it depended on the organization of material at input. Such organization was effective for and positively correlated with categorically associated sets, but it was negatively correlated with alphabetically ordered sets. As a result of this manipulation, Lauer showed that the uncued-recall performance was higher than the cued-recall when the output cues were inconsistent with input organization.

Roediger and Payne (1983) also aimed to show that the uncued-recall performance could be superior to the cued recall. Roediger and Payne, (Experiment 1) who used homographs and category membership found that when the cues were incongruous to the meaning of the targets studied, the uncued-recall memory performance was higher than the cued recall of the same targets. This finding was also gathered by Roediger and Payne (Experiment 2), who reduced the number of categories from four categories down to two categories. In short, the cued-recall performance was only found higher than the uncued recall with congruous cues (Roediger & Payne, 1983). Roediger and Adelson (1980) used the same methodology but allowed the participants to report both cues and targets, Roediger and Adelson (Experiment 3) found the same result as of Roediger and Payne (1983) that the uncued-recall memory performance was higher than the cued-recall performance when the cues were incongruous. These results underlined the importance of semantic encoding, such as the cues produce much better recall of the targets when they are congruous with the sense of the targets compared to when they do not (Roediger & Adelson, 1980).

Some exceptions to recognition failure phenomenon have also been shown with regards to the type of material used. For instance, the recognition failure results could not be gathered when study lists compose of abstract target words rather than concrete ones (e.g., De Vito, 1975; see also Salzberg, 1975; Schwartz, 1975, Epstein, Dupree, & Gronikowski, 1979) as well as when they involve digits as study materials (Gardiner & Tulving, 1980). Besides, the type of cues having various levels of category *set sizes*, refers to the number of available candidates in a category (e.g., large vs. small), yielded that larger sets were less likely to improve recall performance than did smaller set sizes (Nelson & McEvoy, 1979; see also Nelson, McEvoy, & Shreiber, 1990; McEvoy, Nelson, Holley, & Stelnicki, 1992). In regard to the type of extra-list cues, McEvoy and Frederick (1982) showed that 'the degree of control over accessing the domain of information in memory' was also an important factor to observe the fact that extra-list cues could also be effective: A good control over domain search (with no switch between the type of cues encoded at study and test) was seen when all test cues were related to their targets within the same domain.

1.5.2. *The importance of encoding-retrieval match in remembering: Is this a 'myth' as James Nairne thinks?*

In respect to encoding specificity principle and cue effectiveness, the theoretical arguments of James Nairne (2002) seems noteworthy. On encoding specificity principle (in other words, considering the encoding-retrieval match as the most pivotal factor in an effective remembering), Nairne (2002) proposed that the importance of the match between cue and target for recall performance is a ‘myth’. Nairne (2002) disserts his ideas as follows:

> When we remember, we use the information at hand, in the form of retrieval cues, to make a decision about what occurred in the past. But the decision is unlikely to be based on a passive matching process, at least in the majority of retrieval contexts. Remembering is better characterized as an active process of discrimination: We use cues to pick and choose from among viable retrieval candidates. Increasing the encoding–retrieval match generally improves performance, but only because it increases the probability that distinctive features…will come into play. Match, by itself, is not the operative factor behind retention... If a cue is associated to many things, or has been encoded as a part of many trace complexes, then it becomes harder for that cue to elicit any single target trace (Earhard, 1967; Watkins & Watkins, 1975). (p. 390)

Nairne suggested that it would be expected that the performance of recall would enhance with an increase in the amount of similarity between the environment during retrieval and the encoding environment. He explains his proposal with an example. For instance, if an event, i.e., E1, consists of the encoded features of such as X1, X2, and X3, providing participants with only one or two cues (e.g., X1 and X2, or X3 and X1 etc.) could produce decreased performance than providing the participant with all three of these cues (X1, X2, and X3). However, the *cue overload* (the lessened distinctiveness of a particular cue) gets involved into the circumstance in which the performance of recall might be exacerbated. For instance, consider that several features such as X1 and X2 are the encoded features of an event E1, and the feature X1 is unique to an event E1. However, if the feature X2 is also present in other events such as in E2, E3, and E4, providing the participant with X1 and X2 at test, the functional encoding-retrieval match increases (relative to the condition where participant is only provided one of the features), but the memory performance does not necessarily. Because, Nairne proposes that the feature X2 is consistent with other target traces of events as well (e.g., not only in E1, but also in E2, E3, and E4), and so this feature is not distinctive. In short, the cue(s) should be distinctive enough to be effective in remembering the sought-after information correctly (Naire, 2002).

It should be underlined that the idea of ‘cue overload’ of Nairne (2002) together with the reasoning of Higham & Tam (2005; also see Higham 2002), who dealt with recognition failure experiments in terms of type-2 signal detection measurement strategies (see section 6.1.2. for details), became one of the central points to let me drive the main research question followed in this thesis: Providing that the generated candidates are related to each other (e.g., categorically or semantically, or both), the memory quantity could be facilitated, whereas such relatedness between target material should jeopardize the recognition ability of the participants to detect the correct, but inter-related, candidates generated.

The next section will review the specific research questions of the present thesis in detail, which will be followed by laying out the overview of the six experimental studies.

1.6. The research questions

Soon after it was proposed, the generate-recognize theory has been criticised much particularly by the recognition failure of the recallable words phenomenon designed by Thomson and Tulving (1970). The results of the experiments which manipulated various variables in the recognition failure experiments (e.g., type of materials, encoding and retrieval time manipulations) in general favoured the fact that associative continuity hypothesis on which original generate-recognize theory of recall is based does not seem valid. However, Jacoby and Hollingshead (1990) proposed a radical shift on the basic assumption of the theory. They specifically showed that the group of participants who generated responses to ‘strong cues’ had significantly greater memory performance after they were read the target words at study than the group of participants who studied the target words by solving anagram (e.g., the second and the forth letters of the five-letter target words were swooped and the participants reported the words out loud after finding what the original words were) and generated semantic associates to the very same strong cues at test. Hence, Jacoby and Hollingshead showed that the source to generate candidates is not simply semantic memory in which automatic spread of activations takes place (cf. Collins & Loftus, 1975), it rather bases on episodic memory, by which people utilise specific episodic (temporal) information to generate the best candidates. After this achievement, Higham (2002) also showed that the encoding specificity principle that was termed after recognition failure experiments could be valid only when free-report option is adopted and that the observed generation failure could be attenuated by some other variables (e.g., guiding participants on the relationship between the cue presented at test and the target word studied –strong vs.

weak-, Higham & Tam, 2005, or reinstating the semantic association level between study list and the test list, Higham & Tam, 2006).

Therefore, the current research, which aims to understand memory together with metamemory performance in cued and uncued recall of paired associates, considers that the generate-recognize theory could be a fruitful tool to understand memory as well as metamemory processes at recall. Such importance was not only because it has historically become a functional alternative to the idea that recall and recognition are single processes as proposed by *threshold theory* (Postman, 1963; Mandler, Pearlstone, & Koopsman, 1969) and critically discarded by the *frequency effect* (frequent words are recalled better than they are recognized and the infrequent words are recognize better than they are recalled; see e.g., Gorman, 1961; May & Tryk, 1970), but also because it gives much more space to understand metacognitive process instead of direct access view (single-route), which is the prominent counterpart of generate-recognize approach (dual-route). Furthermore, the current research aims to investigate the effect of organization of memory (e.g., categorization), which is a well-grounded memory enhancement method of memory, not only on memory but also metacognitive monitoring performance as its effects on metamemory are not as clear as on memory performance. As will be much clear in the review of second question asked (section 1.6.2.), generate-recognize approach in recall seems to suit much in understanding the possible effects of categorization not only on memory but also on metacognitive monitoring performance.

Hence, the six experimental studies which are reported in this thesis mainly tested the effects of report option (Experiments 1 and 2) and the inter-target association, which was expected to organize study material through categorization (Experiments 1 to 6), on memory and metamemory performance in cued and uncued recall of paired associates. The following sections will review two main research questions held in the present research.

1.6.1. *Does report option matter much in observing what is truly retrieved from memory?*

We do not always have intact memories in all situations. For instance, in responding to a question, or when telling a person an event, one might not only have correct responses or remember the true details about the event, but might also have some incorrect responses or think that the details are correct, which in fact may not be the truth at all. However, if people have an option to give only those responses they are confident, they might be more accurate in their responses compared to the conditions in which they respond with everything they remember (e.g., see Koriat & Goldsmith, 1996c). The researchers dealing with signal detection theory (e.g., Green & Swets, 1966; Klatzky, 1975) suggested that this is more related to the differential level of confidences or strength of the information. As a type-2 SDT concept, however, the term *report criterion* was coined to understand the number and the nature of responses (correct or incorrect) given by the participants (e.g., Koriat & Goldsmith, 1996c); see Figure 2 and Figure 4. Report criterion, as it implies, refers to the criterion people set so as to decide which information is to be reported or withheld in a particular situation. As a result, amongst the responses that are assumed to spread throughout a continuum ranging between having very low and very high confidence levels, people are assumed to report only those items or information being above the report criterion level and do not report the ones below this criterion (Klatzky, 1975); see e.g., Figure 4 for a type-2 SDT modelling.

What does happen when the report criterion is reduced or completely eliminated? In other words, can more correct responses be yielded under forced-report conditions compared to free-report conditions? Evy Cofer (1967), for instance, gave one of the first empirical answers to that question. In his study, the participants learned a list involving 15 items. After a free-recall[2] period, participants were asked to produce enough responses to match the list length even if they needed to make guesses. The results showed that participants produced many responses, however, 'only a few of these responses' produced after the free-recall phase were correct. Therefore, he concluded that participants exhaust the list items that are in storage in free recall, which in a way forces participants to report everything that they remember, and so forced report does not have an incremental effect on obtaining more correct responses (Cofer, 1967).

Investigating *hypermnesia*, Roediger and Payne (1982) also provided some converging evidence to the findings of Cofer (1967). Being a quite interesting phenomenon, hypermnesia refers to the improvement in retrieval performance across repeated tests, which was traditionally measured in terms of the absolute increase in recall across tests (Erdelyi & Becker, 1974). In this sense, hypermnesia might be considered the reverse of forgetting. The standard test to measure this phenomenon

[2] Free recall here refers to the testing conditions in which participants are to free to withhold responses so that they do not have to give same number of responses as the items studied. In this sense, it refers to 'free report'.

comprises of a study phase where participants learn such as pictures or words proceeding with three successive recall tests, which have the duration of seven minutes each (Erdelyi & Becker, 1974; Erdelyi, Buschke & Finskelstein, 1977). Being forerunners of this phenomenon, Erdelyi and Becker (1974) hypothesized that visual coding was the critical factor producing hypermnesia. However, Roediger and Payne (1982; see also Roediger & Thorpe, 1978) argued that this effect could be obtained with the manipulations of recall practice, recall time, or level of recall within a hypermnesia experiment. Hence, Roediger and Payne (1982) argued that this phenomenon is related to the level of recall and it is irrespective of the coding format of the materials.

Along with investigating under which circumstances hypermnesia could be obtained, Roediger and Payne (1985) also compared the recall performance of participants in three successive tests (i.e., Test-1, Test-2 and Test-3) after asking participants to study a list of 70 words under various recall-criterion conditions. The differential recall criterion (report criterion) conditions which were manipulated between participants were: free-recall, uninhibited-recall, and forced-recall conditions. In the *free-recall* condition, participants were instructed to remember as many words as they could in any order, but they were warned not to make guesses. On the other hand, *uninhibited-recall* group was asked to recall as many list words as possible, but they were additionally told to write down any other words which came to mind whilst they were attempting recall. In the *forced-recall* condition, however, participants were given similar instructions and they were told that they had to write down 50 items on their report sheets, even if they needed to make guesses to complete the empty spaces on their report sheets. Therefore, these three groups were set in order to yield the highest report criterion in the free-recall group, the medium level of report criterion in the uninhibited-recall group, and the lowest (no) report criterion in the forced-recall group.

The results of Roediger and Payne (1985) which compared the three groups that were mentioned above and were supposed to vary in terms of recall criterion level (measured by the number of intrusions) yielded the existence of hypernmesia in each recall criterion condition (measured via the difference between Test-1 and Test-3 in terms of percentage recall). More importantly, however, there was not a significant difference in terms of correct responses (e.g., 'hits') between the groups: whereas the free-recall, uninhibited-recall, and forced-recall groups had 2.47, 9.60, and 23.84 intrusions respectively, they did not differ in terms of performance on the third test (28.20, 28.40, 27.42, respectively). Put differently, reducing the report criterion did not have an effect on increasing the overall correct responses, but incorrect ones. Therefore, Roediger and Payne (1985) reasoned the following:

> Contrary to at least the spirit of generate/recognize theories, subjects apparently are not normally generating correct candidate responses in free recall that they reject on an implicit recognition test as being correct, since encouraging them to spew out produces no more correct responses than are found in standard free recall. (p. 6)

Erdelyi, Finks, and Feigin-Pfau (1989) replicated the study of Roediger and Payne (1985) and found converging evidence. In Experiments 1 and 2, Erdelyi et al. found that the forced-recall group did not produce more correct recalls than the free-recall group, even though forced recall produced substantially more false alarms than did free recall. However, in Experiments 3 and 4, they modified the testing procedure and the type of materials utilised in their previous experiments. For instance, they used pictures or concrete words (Experiments 3 & 4) instead of abstract words (Experiments 1 & 2). Also, Erdelyi and his colleagues also asked their participants to free report first and after the participants exhausted all of the responses, they were asked to draw a line below the last response given. Then, the participants were instructed to continue by making guesses until they complete the number of empty spaces provided. They utilised this procedure because they thought there might be a *processing bias* happening, termed by Erdelyi (e.g., 1985). The processing bias, herein, meant that the forced-recall participants might have put less effort into retrieving the items after they were instructed to fill in all of the empty spaces or at least they divide their total effort across all responses to compensate a relative detriment in attention resulted by the earlier responses. In the results of Experiments 3 and 4, they found significantly higher number of correct responses in the forced-report conditions compared to the free-report condition. Roediger, Srivinas, and Waddil (1989), however, commented on the results of Erdelyi et al. (1989) and highlighted that even *large manipulations* on recall criteria produced *only small effects* on the amount of information recalled. Roediger, Srivinas and Waddil (1989) scrutinized the data of Erdelyi et al. (1989) further and showed that their participants made ‘only one item correct for every 10 guesses or intrusions’ on the average when they compared the ratio of the difference between correct responses out of intrusions.

The effect of variations in report criterion on retrieval was also investigated in hypnosis. For instance, Dywan and Bowers (1983) investigating hypnotic hypermnesia, showed 60 slides of line-drawing pictures to 54 participants and these participants were tested in an immediate forced-recall test. The participants were again tested one week later, either under hypnosis or not. The results revealed that hypnosis did not have a facilitative effect on gathering more new correct responses. Furthermore, even though there appeared a few new correct items (1.4 new items), it was accompanied by a high number of errors. Therefore, Dyway and Bowers concluded that the probability of an item to be correctly recalled under hypnosis seemed directly related to the number of items the participants were willing to report. In other words, hypnosis might have only shifted the report criterion, but did not make the memories more accessible. The researchers, however, proposed another possibility: If hypnosis enhanced the vividness of mental imagery, the possibilities generated might have been compelling. As a result, enhanced vividness could have led to a false sense of recognition amongst the hypnotized subjects (Dywan & Bowers, 1983), which implies that there existed a distribution shift rather than a criterion shift (e.g., see Figure 2).

Alongside Dywan and Bowers (1983), Klatzky and Erdelyi (1985) commented on the possible effects of hypnosis on memory from the perspective of signal detection theory. They particularly wished to shed some more light as to whether hypnosis had an effect on the accessibility of information in memory (e.g., *d'*) or it changes the report criterion (e.g., β). They thought hypnosis might be conducive to both components and suggested that the researchers should compare hit and false alarm rates together since focusing on only one of the rates might be misleading or might not be informative at all. Nonetheless, Klatzky and Erdelyi considered that hypnosis might increase correct responses (hits) at a cost of increasing intrusions (false alarms) as well particularly in the experiments where response criterion is not controlled (e.g., forced-report). Therefore, after reviewing the results of the studies pertaining to hypnosis, Klatzky and Erdelyi concluded that hypnosis does seem to neither enhance nor distort memory, but it does seem to vary the report criterion.

Hitherto, the above-mentioned studies utilised single-item study lists. Again, in support of the findings that recall criterion does not seem to have an effect on overall correct recall, some –although a few- experiments using paired-associate learning paradigm showed this same result. For instance, Lockhart (1969) focused on investigating whether *retrieval asymmetry* occurs because of the report option changes.

This phenomenon refers to the conditions in which if word pairs are both high-frequency words and are assumed to be readily available, then the situation is expected to be a favourable one for the demonstration of associative symmetry -where backward and forward semantic association between two words (cue-target) are not different. Hence, it was speculated that correct recall should be independent of the particular item given 'as a cue', since any word in the pair would be helping retrieval equally well (Lockhart, 1969). The results of Lockhart did not reveal any difference between forced-report and free-report conditions (when memory performance is collapsed in terms of the type of cue given: adjective or concrete noun) in a paired-associate learning task. The results mentioned so far might be considered as a suggestion to abandon utilising the forced-report option in memory experiments. More recently, however, some researchers (e.g., Koriat & Goldsmith, 1996a, 1996c; Higham, 2002, 2011; Higham & Tam, 2005, 2006) suggested that it is not necessarily the case since the report criterion differentiations are suggested to be quite informative in the investigation of both memory performance and the metacognitive processes. For instance, Koriat and Goldsmith (1996c) termed *report criterion* (e.g., *Prc*; see Figure 2) to be used in memory contexts as a way of understanding strategic regulation of memory accuracy. They suggested that in a free-report option, not only quantity memory performance of the participants, which has been utilised as a traditional measurement of memory performance in laboratory settings, but also accuracy memory performance should be considered (see Figure 2 for more details on the framework of Koriat & Goldsmith, 1996c). Further, Higham (2002) investigated a highly credited principle in the literature known as the encoding specificity hypothesis of Thomson and Tulving (1970) using a forced-report option as well. Higham (2002) found that whereas weak-cues facilitated the target retrieval compared to no-cue conditions, strong cues (being extra-list) did not facilitate retrieval under 'free-report options'. However, weak and strong cues facilitated the target retrieval compared to the no-cue condition under 'forced report'. More importantly, weak and strong cue conditions did not differ with regards to this facilitative effect. Higham and Tam (2005) replicated the previous experiments of Higham (2002) with several different manipulations such as guiding the participants to make their memory search in the appropriate search sets or domains. Higham and Tam (2005) found the evidence that the reason behind the recognition failure of recallable words phenomenon (Thomson & Tulving, 1970) was in fact due to the failure in the generation process -observed in the inability of extra-list strong cues to generate the possible correct candidates- rather than a failure in recognition. In short, Higham

(2002; see also Higham & Tam, 2005, 2006) showed that report option is a quite important factor to be considered in memory experiments. In other words, these researchers pointed out that when a well-documented principle in the memory literature such as encoding specificity (Thomson & Tulving, 1970) was investigated under forced-report options and with different manipulations such as in the study lists and in the testing procedure, what might actually be happening could be shown much more clearly.

1.6.2. *Do organizational effects on memory have the same facilitative effect on metamemory performance?*

A German psychologist Hermann Ebbinghaus (1885), who tested learning performance of his own in a series of experiments and created serial learning as well as forgetting curves by utilising non-sense syllables, has been considered the pioneer of memory research, particularly in verbal memory research (Crowder & Greene, 2000).

Ebbinghaus (1885) used nonsense syllables, so called because they are pronounceable but meaningless (i.e., 'VOP', 'TUV', etc.). He devised this type of study material in order to have homogenous study materials by which he wished to control the effects of semantic formations that could have been induced by sensible words (such as, 'POT', 'MAY', etc.). Since his leading research, many researchers investigating memory have heavily used the list-learning paradigm. The main reason behind the utilization of this paradigm seems that it provides experimenters with the ability to observe memory performance of people in controlled laboratory settings. Following his research, a huge number of manipulations have also been introduced in the experimental designs to investigate retrieval from memory (e.g., either in terms of recognition, recall, or latency judgements). For instance, researchers have introduced some experimental manipulations such as on the nature of the materials, their mode of study (or presentation), expectations about memory test, the type of these tests, etc. (Crowder & Greene, 2000). In short, since Ebbinghaus (1885), several variables could have been isolated and investigated empirically, and so a huge understanding about human memory has been accumulated (Tulving & Craik, 2000).

However, in relation to the current investigation, amongst many other manipulations introduced in memory research, the effects of *organization* on memory have been of great interest of many researchers. Organization in memory mainly referred to the groupings together of the study list items into larger units (e.g., clusters or chunks), which are usually based on subjectively-decided or experimentally-designated meaningful relationships between the study list items (Brown & Craik, 2000). This behaviour has been observed with the tendency of people to recall the information (e.g., a list of words presented) in an order different than the order in which the information was studied (presented) originally. It should herein be noted that the investigation of organization has been particularly observed in free recall of single-item study lists (Brown, Conover, Flores, & Goodman, 1991).

The effects of organization on memory performance have been investigated heavily. For instance, it has been well documented that semantic organization of study materials (i.e., the study lists involving exemplars of experimentally-constructed or pre-designated particular categories) has a facilitative effect on retrieval (Tulving & Pearlstone, 1966; Cofer, Bruce, & Reicher, 1996; Runquist, 1970; Thomson, Hamlin, & Roenker, 1972), although there have been some exceptions (e.g., Puff, 1970). Tulving and Pearlstone (1966), for instance, tested the memory performance of their participants after the participants studied experimentally-constructed categorized lists. They manipulated the number of categories (one, two, and four) as well as the length of the lists (12, 24, and 48 words). The participants were either presented with the category names as a retrieval cue at the time of testing or without them. They found that participants reported more words when the category names were present than when the names were absent. This difference was even higher when the lists were longer. In short, they showed that higher number of words was retrieved from categorized lists when the organized lists get longer and with a single experimentally-manipulated cue, such as category name. However, Tulving and Pearlstone (1966) accepted that the latter conclusion did not discard the fact that even free-recall[3] tests of random words (e.g., tests having no retrieval cue, like category name) might also yield higher recall performance, if participants subjectively organize these lists via, for instance, using subjectively-constituted meaningful units. This admittance was supported by empirical findings of Tulving (1962; see also, 1966). Tulving (1962) showed that when participants were asked to study the lists composed of 16 unrelated words and the participants were tested under repeated tests after different orders of the list presentation, they seemed to impose a sequential structure on their recall and this was

[3] The cued-recall groups of Tulving and Pearlstone (1966) were cued only in terms of the fact that the participants were provided with category names as cues. However, cued recall conventionally referred to those conditions where each to-be-remembered item is primed with a cue (or several cues). In that sense, the cued-recall and free-recall groups of Tulving & Pearlstone (1966) were both free-recall groups, only one of which (the cued-recall one) were provided with cue(s) at testing, such as category names.

increased with repeated exposure and recall of the list. In other words, participants show a sequence in recall of the list words in a stereotyped manner, which are constructed idiosyncratically (Tulving, 1962, 1966).

Despite the fact that a facilitative effect of categorization on retrieval exists, the effects of studying categorized lists on metamemory have remained unclear. Recently, however, Guerin and Miller (2008) reported that they had the first direct comparison of the effects of categorically-organized lists on recall and recognition. They tested the memory performance of three groups of participants; cued recall, uncued (free) recall, and recognition, after they had studied categorized and uncategorized lists. Guerin and Miller found that categorization increased the false alarm rate substantially although it did not have any effect on hit rate in recognition performance. However, under both cued and uncued recall, categorized lists resulted in higher hit rates than uncategorized lists despite yielding not significant changes in false alarm rates. In terms of the recognition results, the researchers found higher sensitivity (e.g., d_a; the discrimination index used when standard deviations of two distributions are not assumed equal, e.g., see Macmillan & Creelman, 1991) for uncategorized lists than categorized lists. Therefore, they concluded that categorization impaired recognition performance, although recall was superior for categorized lists (Guerin & Miller, 2008). However, they accepted that they made only some claims about the 'relative direction' pertaining to the effects of organization on recall and recognition, since direct comparison of these two processes raise difficulties in the interpretation of results. Because recognition performance might not be linearly related to recall performance, Guerin and Miller indeed proposed 'a possible reason' only for the impairment of recognition with the facilitation of recall when the lists were categorized. They reasoned that organized lists share a common feature, thus, items become less novel and less distinctive so that it could be expected that the items in organized (categorized) lists were recognized poorer (Guerin & Miller, 2008).

In relation to the reasoning of Guerin and Miller (2008) on distinctiveness, Higham and Tam (2005; see also 2006), who found generation failure in the traditional encoding specificity principle experiments (Thomson & Tulving, 1970), suggested the following. The failure in generation was shown when cued-recall performance of extra-list strong cues in target production was worse than the performance of control group who were asked to generate responses and not given any study list in advance (Higham & Tam, 2005, 2006). As an argument pertaining to the reason of the generation failure, they concluded that the participants search inappropriate domains in memory. Further, Higham and Tam (2005) proposed that if the cue-target semantic association is strong, generation of possible candidates could be enhanced, but the recognizing the correct candidate (target) would be deteriorated since several interrelated items would be also generated. On the other hand, if the cue-target association is weak, the generated response might either be the only response or the target item generated becomes distinctive enough to be recognized amongst the candidates. In short, the arguments and findings of Higham (2002; Higham & Tam, 2005, 2006) as well as Guerin and Miller (2008) were quite promising to set the ideas and the rationale of the current study, especially regarding the issue of distinctiveness in generated candidates.

1.7. Overview of the experiments

The six experiments that are reported in the following chapters in detail tried to answer two main research questions outlined in the previous section. To be more specific, Experiments 1 and 2 tested the first and the second questions simultaneously. In Experiment 1, the report type was varied between experimental (instructed to 'report only the target words') and the control group participants (instructed to 'report any (all) words studied') who were tested under uncued recall after they had studied mixed list of strong and weak target-cue pairs (e.g., Bulb-Light, Street-Mineral). Following Experiment 1, Experiment 2 manipulated the report-option between free and forced-report groups of cued as well as uncued-recall participants. Experiment 2, however, introduced a unique variable that manipulated the level of categorical relatedness between target items (high vs. low). Experiment 3 as well as Experiments 4, 5, and 6, then, dealt with the second research question further, which was asked to investigate organization effects on memory and metamemory performance (e.g., type-2 signal detection *d'*). It specifically aimed to scrutinize the effects of encoding (focus on the pairs: target-focused vs. pair-focused, in which participants are informed about the places of cues and target or not) as well as the retrieval time manipulations (report instruction: constrained cued recall vs. liberated cued recall, through which the way of measurement strategy regarding the correctness of the responses given was revealed differently) on memory and metamemory performance. Experiment 4 critically manipulated the number of categories to which target items belonged (two vs. six vs. twenty-four) so as to show a clear dissociative pattern between memory and metamemory performance due to high inter-target association. After substantiating a clear dissociation, Experiment 5 aimed to diminish the dissociation found in Experiment 4 via manipulating the encoding process (e.g., interactive imagery vs. rote repletion) to achieve individuation of the pairs within the study lists. Lastly, Experiment 6 used homograph words to be targets (the words having different meanings depending on the contexts in which they are used) and aimed to test whether specific episodic information about the contexts the words implied (e.g., the target words implied particular categories with the remaining target words) is more vital to lead the dissociative pattern between memory quantity and monitoring performance than the simple spread of activations between target items.

CHAPTER 2
Experiment 1

2.1. The aims and expectations

The generate-recognize theory and its assumptions were considered as a backbone for the current research, mainly because it gives much more space to investigate metacognitive processes in retrieval, such as in recall, rather than direct retrieval approach (see e.g., Bahrick, 1969, 1970). Hence, the first aim centring on the recall criterion differentiations (varying the report option) in the present study was to investigate the effects of report criterion changes both on memory and metamemory performance using paired-associate learning. Specifically, the present study basically aimed to investigate whether reduced recall criterion level would increase overall correct responses in a paired-associate learning experiment. Additionally, it aimed to investigate the effects of semantic association between pairs on different types of memory performance (memory quantity and memory accuracy) as well as on metacognitive processes, particularly monitoring. The rationale and the details of the experimental design are as follows.

First, the present study intended to investigate as to whether reduced recall criterion level yields more correct responses in a paired-associate learning task of recall. The *paired-associate learning paradigm*, which was invented by Calkins (1894), traditionally refers to the procedure involving the pairing of two items such as the words being a stimulus (cue) and a response (target). The learner (participant) is provided with the stimulus and he/she is asked to respond with the response item paired with the stimulus item. The numbers of stimulus and response items might vary. In the current experiment, the level of the report criterion was varied between an *experimental group* and a *control group*, both of which were tested under free-report and uncued recall[4]. In order to investigate the effect of report criterion on memory in uncued recall, where participants are not provided with any cues at the time of testing, the experimental-group participants were asked to report 'only target words' (e.g., words on the left hand side of the pairs) and the control-group participants were asked to report 'as many words as possible (any words) presented'.

[4] 'Free recall' term will be used as 'uncued recall' in this experiment and in the following ones since report options will be defined as 'free report' and 'forced report'; see e.g., Higham & Guzel, 2011 for definitions of 'cued recall' and 'cueing'.

One might ask why the control group would have lower recall criterion than the experimental group. The idea behind this manipulation could be explained as follows. Reduced recall criterion in the current experiment, which is aimed to be achieved with the control group, mostly resembles the 'uninhibited free-recall condition' used by Roediger and Payne (1985). In their study, participants in the uninhibited-recall condition were asked to recall as many list items as possible in any order. Additionally, they were instructed to write down any other words that came to mind whilst recalling the list items, even though they knew that they were only guessing. Although participants in the current study were not asked to make uninhibited guesses, reporting any words for the control-group participants was thought to yield relatively lower recall criterion than the experimental condition, in which participants were required to inhibit reporting any other words (e.g., cues) remembered. In other words, the control-group participants could both report the predetermined targets (e.g., words on the left-hand side in the pairs) and they did not have to inhibit reporting particular words (e.g., the words on the right-hand side in the pairs, which are indeed the cue words for the experimental group). Additionally, the control group did not have a necessity to remember the places of the particular words to be remembered as to whether the remembered word is a cue or a target. Hence, the experimental-group participants are expected to pay more attention to the placing of words and inhibit reporting of the incorrect responses (cue words) at the time of retrieval, which thereby is expected to result in a stricter recall criterion than the control-group participants (see section 2.2.5. for more details about the procedure).

Second, semantic association (SA) level between target and cue in a paired-associate learning paradigm was manipulated as high vs. low. The main reason behind this manipulation based on two objectives. The *first* objective was to have facilitated recall by high semantic association between cue and target. Based on the previous findings (e.g., see Tulving & Pearlstone, 1966; Thomson & Tulving, 1970; Roediger & Adelson, 1980; Roediger & Payne, 1982), it was anticipated that this kind of aid would improve memory performance of the participants, compared to those targets having less (or no) semantically associated cues. However, the *second* objective is more central. Although the effects of semantic association on memory are well-known, its effects on metamemory are unclear. Hence, the study aimed to investigate both memory and metamemory performance (such as monitoring) at the same time in recall. Although there has appeared very little suggestions in the literature pertaining to the effects of manipulations made in the study and/or retrieval contexts on 'both' memory and metamemory performance, the ideas and findings of Higham and Tam (2005; see also 2006), for instance, are quite promising in that sense. That is, Higham and Tam (2005) suggested the following in the generation failure experiments with regards to the effects of semantic association level between cue and target items on memory and monitoring performance (Higham & Tam, 2005, 2006; see also Higham, 2002). Should the cues be strong associates of target (e.g., forward semantic association from cue to target), then the generation of possible candidates being a target is facilitated. However, the strong association between cue and target would also facilitate the process of generating several 'interrelated' items that thereby could result in the deterioration of recognition process. On the other hand, weak associate cues were supposed to restrict the search set size via, for instance, the instruction to avoid generating related items in target recall or, even if they do not avoid to do so, generated items will less likely to be related to the target word. As a result, just like Higham and Tam (2005) reasoned "the target may be the only candidate to be considered seriously as a response to each (reinstated) weak cue, either because it was the only generated candidate, or because it 'stood out from the crowd' being unrelated to the other items in the candidate set" (p. 607).

Considering the proposals of Higham and Tam (2005, 2006) with regards to the possible effects of semantic association on generation and recognition processes, the semantic association between cue and target pairs in the current experiment was bidirectional. In other words, the forward and backward semantic associations between cue and target words either existed at the same time or no semantic association between the words in the pairs existed at all. This manipulation resulted in two levels of SA manipulation: high-SA vs. low-SA target pairs. It was reasoned that should the basic assumption of generate-recognize theories' proposing two stages in recall -generating possible candidates first and then recognizing the target from amongst the candidates generated- is valid, beside yielding higher correct responses in reduced-report criterion (in control group), monitoring performance in high-SA target recall would be expected to be lower compared to low-SA target recall. Because, cue and targets would be bidirectionally related to each other in terms of semantic association (high-SA pairs), which thereby was expected to facilitate not only target retrieval but also cue retrieval. As a result, the facilitated cue retrieval would lower monitoring performance in high-SA target recall or at least would render the monitoring performance in high-SA target recall be comparable to low-SA targets, although memory quantity was expected to be higher in high-SA target recall compared to low-SA target recall. In short, the facilitative effect of semantic association on memory performance is not expected in monitoring performance.

In order to reach the objectives mentioned above, the participants in the experimental and control groups studied mixed lists of paired words (e.g., target-cue), in which half of the pairs are semantically associated to each other (high-SA pairs) and the remaining pairs are not semantically associated (low-SA pairs). The participants were given a free-report option at test and were tested in uncued recall. The difference between groups was solely at the time of testing where participants were either asked to report the targets only (the experimental group) or to report any of the words from the studied lists (the control group). In order to measure the retrieval (memory) and metamemory performance at the same time, memory quantity and memory accuracy (Koriat & Goldsmith, e.g., 1996c) as well as the promising measurement strategies of signal detection theory, which is particularly based on type-2 signal detection measurements were utilised (see section 1.3. for further details about the approaches).

Therefore, it was hypothesized that;

(a) the number of correct responses would be higher when recall criterion is reduced. In other words, the control group is expected to have higher memory quantity (MQ) than the experimental group. Such a result would be able to confirm the validity of generate-recognize theories, which seemed to be undermined by failing to find facilitative effect of reduced report criterion on yielding more correct responses, particularly in the studies using single-item study lists (e.g., Dywan & Bowers, 1983; Erdelyi et al., 1989). Beside the fact that participants would study the pairs without being informed which side of the pairs would be the to-be-remember items (targets) in advance, the bidirectional semantic association between high-SA pairs was expected to be a hindrance rather than a facilitative effect to report 'only' the target words in the experimental group.

(b) Therefore, the existence of semantic association between cues and targets would have an incremental effect in terms of target recall -even though participants recall targets without cues (uncued recall).

(c) However, the participants were expected to have higher memory accuracy (MA) than memory quantity as a reflection that they effectively and strategically regulate their memory accuracy. More importantly, participants are expected to have higher memory quantity for high-SA targets than low-SA target recall, but their memory accuracy would not differ from each other since high-SA target recall was expected to facilitate generation process of the targets, but some of the reported items would be incorrect responses (particularly cues). Therefore, as the number of the reported items would increase in high-SA target recall (denominator of MA calculation) and this number would be lower in low-SA target recall, the proportion of the number of correct responses out of total items reported (MA) would be comparable in high-SA target recall and low-SA target recall. Complimentary to the comparisons between MQ and MA, monitoring performance (e.g., area under the curve scores; see section 2.13. for the details) in high-SA target recall was expected to be lower than (at least comparable to) in low-SA target recall, since the participants were expected to yield higher number of incorrect responses reported in high-SA target recall as a result of facilitated generation process that result in a difficulty in recognizing the correct response amongst the generated candidates.

2.2. *Method*

2.2.1. *Participants*

Forty-eight undergraduate or graduate psychology students (age: *M* = 20.81, *SD* = 3.54) in the University of Southampton, England, whose first language was English, participated in the experiment. Each participant was compensated for his/her time with course credits or £5 cash payment. Thirteen of the participants (27.1%) were male and 35 (72.9%) of them were female. Each of the participants was randomly assigned to one of the conditions of the experiment: the experimental condition (*n* = 24; age: *M* = 20.46, *SD* = 2.65) and the control condition (*n* = 24; age: *M* = 21.17, *SD* = 4.28).

2.2.2. *Materials*

2.2.3. *Study lists*

A list of 30 word pairs was constructed for the study. In each pair, one word was the target word and the other one was the cue word (e.g., Target - Cue). The list consisted of 15 word pairs having high semantic association together with 15 word pairs having low semantic association so that the list was a mixed list. The SA variable was manipulated according to the backward and forward association levels between each word in the pairs based on the University of South Florida Free Association Norms (Nelson, McEvoy & Schreiber, 1998). High-SA word pairs (i.e., BULB - LIGHT) had both strong forward semantic association (association from target to cue) and backward semantic association (association from cue to target). On the other hand, low-SA word pairs (i.e., ROSE - JACKET) lacked such an association. The SA level in each pair was counterbalanced, and so there appeared two versions of the list (List-1 and List-2, each of which had 15 high-SA and 15 low-SA target pairs); see Appendix A. That is, each target word was paired with its high associate cue in List-1 (e.g., BULB - LIGHT) if it is paired with its low associate cue in List-2 (BULB - PIANO), and each target was paired with its low associate cue in List-1 (e.g., CALCULUS - MILITARY) if it is paired with its high associate in List-2 (CALCULUS - MATH). Therefore, two separate lists, involving the same 30 target words but paired with 30 different cue words in each, were constructed. In short, each target was paired with each level of semantically associated cue word (high or low) between lists. Based on the norms study of Nelson et al. (1998), the forward and backward associations in List-1 were: $M = 50.67$, $SD = 21.70$; and, $M = 53.33$, $SD = 21.18$, respectively. In List-2, the forward association was $M = 49.4$, $SD = 19.54$ and the backward association was $M = 50.8$, $SD = 24.97$. The first and second lists did not differ significantly in terms of their forward and backward association levels, $t(14) = .764$, $p > .05$; $t(14) = .874$, $p > .05$, respectively. For the sake of not having a confounding effect of semantic association, the lists were arranged in a way that none of the cues were semantically associated to any other cues, none of the targets were semantically associated to any other targets, and none of the cues were semantically associated to any other targets (except their very own targets if the pair is a high-SA pair); see Appendix A for the study lists created for Experiment 1.

2.2.4. *Experimental design*

The experiment used a 2(semantic association level of pairs: high vs. low) X 2(reporting task: reporting only targets vs. reporting all words) mixed design, with semantic association level of pairs manipulated within participants and reporting task manipulated between participants. The dependent variables were confidence levels and memory performance, measured in terms of memory quantity and memory accuracy. As a type-2 signal detection measurement strategy, the metacognitive monitoring performance of the participants was measured with Area Under a Curve (AUC) calculations (see section 2.3.4. for the detailed explanation of AUC calculations). Memory quantity (MQ) was defined as the percentage of correctly recalled targets out of the total number of targets. On the other hand, memory accuracy (MA) was defined as the percentage of correctly recalled targets out of the total number of the words reported by the participant. Participants were asked to recall and report 'only the target words' in the experimental condition and they were asked to recall and report 'as many words as possible (all words)' in the control condition. Therefore, the term 'reporting task' will be used with 'groups' interchangeably.

2.2.5. *Procedure*

Participants were tested individually in a quiet and dimly lit Cognitive Laboratory, located in the School of Psychology at the University of Southampton, England. Each participant read and signed a written informed consent form before the study started. The study involved a study-test cycle: a computerized study and a paper-pencil test. Runtime Revolution 2.5 computer program was utilised for the study phase by which each participant was presented with a total of 30 word pairs on a Macintosh computer. Fifteen of the pairs were high-SA pairs and 15 of them were low-SA pairs. Since the lists were counterbalanced in terms of the pairs' semantic association level, participants were presented with either List-1 or List-2, based on their attendance sequence. Additionally, the order of presentation of the pairs was totally randomized by the computer program. Each pair appeared on the screen for 3 seconds with 1-second inter-stimulus interval (ISI). All of the participants were instructed to attend to all of the words in the list to be presented in advance of the study and instructed that they would be required to try to remember as many words as possible after the presentation. They were not informed in advance about which word was the target and which word was the cue.

In the test phase, the experimental group was instructed to remember and report only the target words, which were 'the words on the left-hand side of the pairs'. The control-group participants were instructed to remember and report 'as many words as possible', regardless of whether they were on the left or the right during study. Each participant was given reporting sheets that contained the instructions of the reporting task specific to the condition, columns to write down the responses, and rows to indicate the confidence ratings for the correctness of the response given. For confidence ratings, each participant was asked to define his/her confidence level for each word he/she reported on how confident he/she was on the correctness of the response. They rated their confidence levels on a Likert-type scale ranging between 1 and 7 (1 = "Not at all confident correct", 4 = "Fairly confident correct", 7 = "Completely confident correct"). Participants had no time restriction during the reporting phase. After the test phase, participants were debriefed with a debriefing form and the researcher responded to their enquiries about the study. The study, which was self-paced at testing, lasted 20-30 minutes (see Appendix A for the informed consent form, reporting sheets, and the debriefing form used in Experiment 1).

2.3. Results[5]

2.3.1. *Memory quantity*

The first analysis was a 2(group: experimental vs. control) X 2(semantic association level in pairs: high vs. low) mixed ANOVA on memory quantity (MQ); see Table 2. The results showed that group main effect was significant, $F(1, 44) = 5.459$, $p = .024$, $\eta^2 = .110$: The control group had higher memory quantity ($M = 31.7$, $SE = 1.9$) than the experimental group ($M = 25.4$, $SE = 1.9$). This result revealed the reduced recall criterion increased the number of correct responses reported. The main effect of SA was also significant, $F(1, 44) = 11.525$, $p = .001$, $\eta^2 = 208$. The participants (the experimental and the control groups together) retrieved higher percentage of high-SA targets ($M = 33.0$, $SE = 1.8$) than low-SA targets ($M = 24.5$. $SE = 2.0$), which indicate the semantic association level existing between the pairs help retrieval of the targets. The results did not reveal any interaction effect between group and SA level, $F<1$.

2.3.2. *Memory accuracy*

First, a 2(group: experimental vs. control) X 2(semantic association level in pairs: high vs. low) mixed ANOVA was conducted on memory accuracy (MA); see Table 2. The group main effect was significant, $F(1, 44) = 6.062$, $p = .018$, $\eta^2 = 121$. The experimental-group participants had significantly higher MA ($M = 58.1$, $SE = 3.7$) compared to the control-group participants ($M = 45.1$, $SE = 3.7$). As expected, however, SA level main effect was not significant, $F(1, 44) = 3.041$, $p = .088$. Further, the results did not show any interaction effect between group and SA level on memory accuracy, $F<1$.

Table 2

Means of Memory Quantity and Memory Accuracy (%) in the Experimental and Control Groups as a Function of Semantic Association Level Between Pairs

	Memory quantity			Memory accuracy		
	Semantic association level between pairs					
Groups	High	Low	High & Low	High	Low	High & Low
Experimental	30 (13)	20 (10)	25 (8)	62 (26)	55 (23)	58 (21)
Control	36 (12)	28 (16)	32 (11)	47 (13)	44 (17)	45 (15)

Note. Standard deviations are in parentheses.

[5] Two cases, one in the experimental group and one in the control group, were detected as being outliers on the basis of their z-scores and Mahalanobis distance. Hence, they were excluded from the analyses.

2.3.3. *Intrusion errors*

Participants reported some incorrect responses as well. These intrusion errors were categorized into two types: *cue words* and *extra-list words*. Extra-list words referred to the words that were neither cues nor targets. The intrusions errors were calculated and reported in terms of number reported rather than percentage, since the denominator for the extra-list words reported was unknown. Table 3 displays the mean confidence levels given for target, cue, and extra-list words as a function of semantic association in pairs (high vs. low) in experimental and control groups.

Cue-type intrusions were inevitably expected as recall was uncued in each group. A 2(group: experimental vs. control) X 2(semantic association in pairs: high vs. low) mixed ANOVA was conducted on the percentage of cue-type intrusion errors. The results indicated a group main effect, $F(1, 44) = 53.179$, $p < .001$, $\eta^2 = .547$. The experimental-group participants had significantly lower number of cue-type intrusions ($M = 1.33$, $SE = .31$) than the control-group participants ($M = 4.48$, $SE = .31$). This result was expected since the control-group participants were instructed to recall and report as many words as possible (any words from the studied list), whereas the experimental-group participants were responsible for reporting only the target words. Additionally, semantic association level main effect was also significant, $F(1, 44) = 5.563$, $p = .023$, $\eta^2 = .112$. Participants had significantly higher percentage of cue-type intrusions for high-SA target recall ($M = 3.30$, $SE = .29$) than for low-SA target recall ($M = 2.50$, $SE = .26$). Lastly, the results did not reveal any interaction effect between group and SA level on the percentage of cue-type intrusions, $F(1, 44) = 1.467$, $p > .05$.

In addition to cues, participants also reported extra-list words as intrusion errors. The presented lists were mixed, involving both high-SA and low-SA pairs so that extra-list words given could not be separated in terms of whether they were the reported items associated with high-SA targets or with low-SA targets. As a result, the number of extra-list words was treated exactly the same for both high and low-SA target recalls. The results showed that the experimental and the control-group participants did not differ in terms of the percentage of extra-list words reported ($M = 1.61$, $SD = .31$; $M = 1.65$, $SD = .38$, respectively), $t(44) = -.089$, $p > .05$.

When the percentage of cues and extra-list words reported by the participants were compared, the results also revealed that the experimental-group participants did not differ in terms of percentage of cues and extra-list words reported for high-SA target recall ($M = 1.52$, $SD = 2.0$; $M = 1.61$, $SD = 1.47$, respectively), $t(23) = -.163$, $p > .05$, as well as for low-SA target recall ($M = 1.13$, $SD = .87$; $M = 1.61$, $SD = 1.47$, respectively), *t*(22) = -1.392, *p*> .05. However, the control-group participants reported significantly more cue words (*M* = 5.09, *SD* = 1.93) than extra-list words (*M* = 1.65, *SD* = 1.82) in high-SA target recall, *t*(22) = 7.819, *p*< .001, as well as in low-SA target recall (*M* = 3.87, *SD* = 2.43; *M* = 1.65. *SD* = 1.82, respectively), *t*(22) = 3.275, *p* = .003.

Table 3

Mean Confidence Levels of Targets, Cues, and Extra-List Words As a Function of Semantic Association Level in Pairs (High vs. Low) and Group (Experimental vs. Control)

	Target		Cue		Extra-list
Groups	High	Low	High	Low	High & Low
Experimental	5.6 (1.2)	6.0 (1.2)	4.5 (1.4)	4.7 (1.9)	3.3 (1.7)
Control	6.6 (0.7)	6.5 (0.7)	6.6 (0.5)	6.5 (0.7)	3.7 (2.0)

Note. The study lists were mixed-lists so that extra-list intrusion errors could not be separated as to whether they were retrieved as a result of the sematic activations from high-SA pairs or low-SA pairs. Therefore, these intrusion errors were counted exactly the same in high and low-SA recalls.

2.3.4. ***Monitoring performance***

Being a nonparametric estimator, area under a curve (AUC) was calculated for each participant in order to measure the monitoring performance of the participants. For the calculation of AUC scores, the hit rates and the false alarm rates of each participant are needed. The hit and false alarm rates were calculated on the basis of a contingency table yielding the number of correctly and incorrectly recalled words throughout *cumulative confidence levels* (1+, 2+, 3+, 4+, 5+, 6+, and 7). The cumulative confidence refers to the confidence levels given for the responses and linked with one of the subjective confidence levels being either the one in consideration (e.g., 2+) and above this particular confidence level (confidence level, 2) up to the highest confidence level available in the rating (e.g., 7) as well as below this confidence level (1). Therefore, ‘2+’ referred to the confidence levels which considers the confidence level ‘2’ as a reference point and counts the correct and incorrect responses given a confidence level either above 2 (2, 3, 4, 5, 6, & 7) and below 2 (1); and ‘3+’ refers to the consideration of correct and incorrect responses given a confidence level ‘3’ and above (3, 4, 5, 6, and 7) and below ‘3’ (1 & 2), and so on.

Table 4

A 2(Type of Response: Correct vs. Incorrect) X 2 (Confidence Level: Above vs. Below the Confidence Level in Consideration) Contingency Table Constructed to Calculate Hit Rates and False Alarm Rates at 2+ Confidence Level

	Type of the response	
Confidence level	Correct	Incorrect
Above (include 2, 3, 4, 5, 6, or 7 confidence level)	a	c
Below (include only confidence level 1)	b	d

Note. a = number of correct words recalled and given a confidence level either 2, 3, 4, 5, 6, or 7; c = number of incorrect words recalled and given a confidence level either 2, 3, 4, 5, 6, or 7; b = number of correct words recalled and given a confidence level 1; d = number of incorrect words recalled and given a confidence level 1.

Based on Table 4, when one wishes to calculate hit rates and false alarm rates at, for instance, 2+ confidence level for each participant, the calculations consider the number of correct and incorrect responses given and linked with one of the subjective confidence levels amongst ‘2 and above’ as well as ‘below 2’. Therefore, as it is seen in Table 4, ‘a’ refers to the number of *correct* words recalled and given one of the confidence levels either ‘2, 3, 4, 5, 6, or 7’; and, ‘b’ refers to the number of *correct* words recalled and given a confidence level ‘1’. On the other hand, ‘c’ refers to the number of *incorrect* words recalled and given one of the confidence levels either ‘2, 3, 4, 5, 6 or 7’; ‘d’ refers to the number of *incorrect* words recalled and given a confidence level ‘1’. Based on the contingency table, Table 4;

$$HR = a / (a + b) \quad (1)$$

$$FAR = c / (c + d) \quad (2)$$

in which HR is the proportion of correct responses out of all possible correct responses, and FAR is the proportion of incorrect responses out of all possible incorrect responses at a particular cumulative confidence level.

As a result, the hit and false alarm rates were calculated for each cumulative confidence level (1+, 2+, 3+, 4+, 5+, 6+, and 7) and they were plotted on ROC (Receiver Operating Characteristics) curves (see Figure 5). The sum of trapezoidal areas between two successive cumulative confidence levels, which are plotted on the ROC curve, yields an indication of monitoring performance (AUC scores) of a particular group[6]. Hence, the higher the AUC score is, the better is the monitoring performance.

In order to analyse monitoring performance of the participants, a 2(group: experimental vs. control) X 2(semantic association level in pairs: high vs. low) mixed ANOVA was conducted on AUC scores; see Table 5. The ROC curves of each group in high-SA and low-SA target recalls are displayed in Figure 5. The results revealed that group main effect was significant, $F(1,44) = 9.669$, $p = .003$, $\eta^2 = .180$. The experimental-group participants monitored their responses better ($M = .78$, $SE = .03$) than the control-group participants ($M = .64$, $SE = .03$). The results did not show any SA level main effect; and, SA level and group interaction effect, both $Fs<1$.

Table 5

Area Under The Curve Scores as a Function of Group and Semantic Association (SA) Level Between Pairs in Recall

	SA level between pairs in recall	
Group	High	Low
Experimental	.76 (.27)	.81 (.21)
Control	.63 (.12)	.65 (.17)

Note. Standard deviations are in parentheses.

The intersection points of hit and false alarm rates at cumulative confidence levels were shown in ROC curves in Figure 5. As it is seen in Figure 5, the experimental group monitored their responses better than the control group both when

[6] When AUC calculation is applied for 1+ confidence level, it yields exactly the same HR and FA rates for each participants, '1.00'. For example, consider the following calculations for 1+ cumulative confidence level. Based on Table 3, 'a' equals to a particular number of correct responses, say, 'X'; and 'b' is always 'zero', since there exists no responses attached a confidence level below 1; and, 'c' equals to a particular number again, say, 'Y'; lastly, 'd' is always 'zero', because there is again no such response linked with a confidence level below 1. Then, based on the formula (1) and (2);

$$HR = X / (X + 0) = 1.0 \quad (3)$$
$$FA = Y / (Y + 0) = 1.0 \quad (4)$$

As a result, AUC score at '1+' is always 'zero' (e.g., see Figure 5).

the cue-target pairs had high semantic association and when they had low semantic association. Since both groups were 'free to report', the participants seemed to report the extra-list words voluntarily so that these responses were attached with fairly high confidence levels, but not with the lowest confidence levels possible (e.g., 1 or 2): The mean confidence levels (MeanCF) for extra-list words reported in experimental group was 3.27, (*SD* =1.65), and MeanCF in the control group for extra-list words reported was 3.64 (*SD* =1.93).

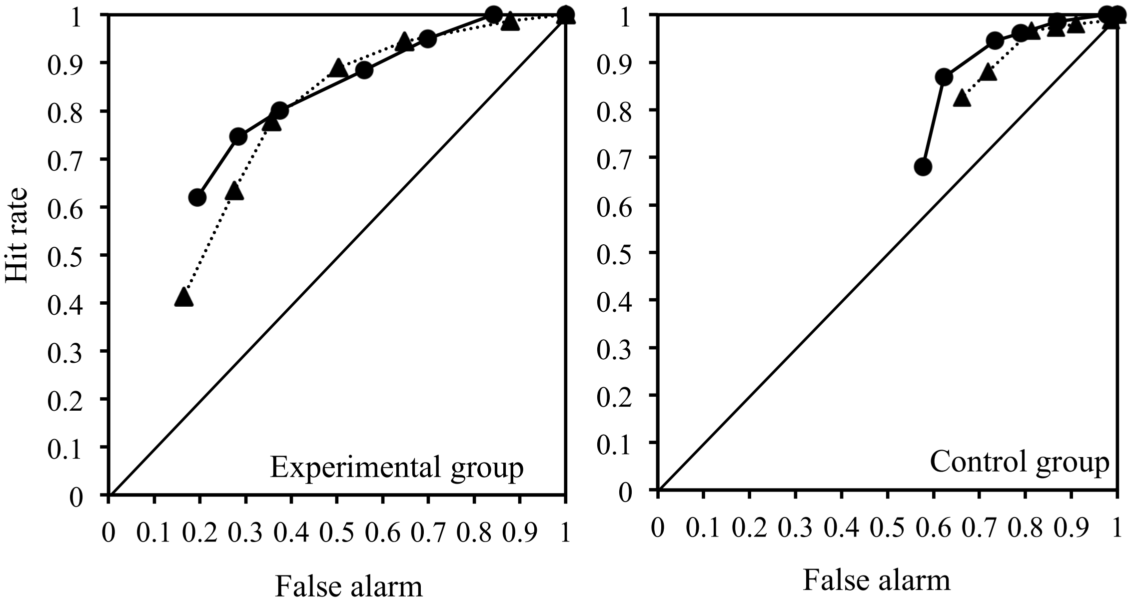


Figure 5. ROC curves displaying the intersection points of hit and false alarm rates at each cumulative confidence level in high-SA and in low-SA target recalls of the experimental and control groups. Whilst straight lines (—) display the ROCs in High-SA target recall, dashed lines (---) display ROCs in Low-SA target recall.

2.4. Discussion

The results on memory quantity confirmed that when participants have stringent report criterion (the experimental group), they had lower correct responses compared to when they have less stringent report criterion (the control group). In other words, as a counter evidence to the previous results in the literature (e.g., Lockhart, 1969; Roediger & Payne, 1985) and confirming the assumptions of generate-recognize approach, the participants indeed had a tendency to withhold some correct responses that they had due to a stringent report criterion.

The study further showed that the existence of semantic association level between pairs was a facilitative effect on retrieval performance (memory quantity).

That is, participants seemed to utilise the semantic association when available to yield better quantity memory performance. This finding supported the previous literature, which suggests that there exists a facilitative effect of semantic association between cues and to-be-remembered items on retrieval (see e.g., Experiments 1 & 3, Thomson & Tulving, 1970; Experiment 4, Wiseman & Tulving, 1976). However, when this facilitative effect was investigated with metacognitive performance in free recall, the following findings were gathered.

Notwithstanding the higher memory quantity found in high-SA target recall than in low-SA target recall, participants demonstrated similar accuracy memory performance in high-SA and low-SA target recalls. In other words, even though the participants recalled fewer correct responses for low-SA target recall (indexed with lower memory quantity) compared to high-SA target recall (M = 24.5; M = 33.0, respectively), they were not less accurate in their performance in low-SA target recall than they were in high-SA target recall (M = 49.5, M = 54.5, respectively) (see Table 2). In short, high semantic association between pairs played a facilitative role in target recall, however, it did not have the same facilitative effect in accuracy.

The monitoring performance of the participants were not only drawn from the comparisons of MQ and MA suggested by Koriat and Goldsmith (1996c), but also by AUC calculations, which is based on type-2 signal detection and takes confidence levels into account. When AUC results were considered, the existence of SA in pairs did *not* have any facilitative effect on monitoring performance, unlike it had on retrieval performance (MQ). That is, the participants (the experimental and the control groups together) had comparable monitoring (AUC) in high and low-SA target recall (regardless of whether cues reported were counted incorrect or cues reported were counted correct for the control group). Parallel to this observation, the experimental-group participants did not differ in terms of monitoring performance between high-SA target recall (M = .76) and low-SA target recall (M = .81), $t(22) = -.744$, $p > .05$. Further, the control-group participants had also comparable monitoring performance between high-ITA target recall (M = .63) and low-SA target recall (M = .65), $t(23) = -.640$, $p > .05$.

In short, counter to the earlier findings (e.g., Lockhart, 1969; Roediger & Payne, 1985), the present study supported the assumptions of generate-recognize models that participants withhold some correct responses that they had, due to a controlled report criterion (e.g., under a relatively more stringent report criterion). Further, Experiment 1 showed that the existence of semantic association yield higher retrieval performance, however, such existence expectedly had such incremental effect ‘only’ on memory performance. In other words, the findings revealed that a facilitative effect on memory driven by the high semantic association level does *not* seem to have the same merit on monitoring performance. This result seems to confirm the expectations based on the assumptions of generate-recognize theory. As argued by Higham and Tam (2005), should a facilitative effect be created on generation that thereby increases the retrieval performance, this facilitation might deteriorate the monitoring performance, or at least, does not have the same positive effect on recognition (reflected on reduced -or at least not facilitated- metacognitive monitoring). As a result, the argument of Higham (2002; see also Higham & Tam, 2005, 2006) was confirmed in the way that the existence of semantic association between pairs could facilitate generation process (observed via higher MQ performance in high-SA target recall compared to low-SA target recall), however it does not have the same facilitative effect on monitoring performance (observed with comparable MA & AUC scores in high-SA and low-SA target recalls). Therefore, this pattern was a critical observation to provide me with a reason to pursue the track that some possibly well-known facilitative effects on memory performance (i.e., semantic association, categorization etc.) might not be effective on metamemory performance.

Hence, on the basis of the findings obtained in the current study, the next study will introduce a new experimental design which manipulated the study list structure in paired-associate learning to investigate the effects of not only semantic but also episodic features of the study lists on memory and metamemory, such as the knowledge gained about the inter-target associations between target items.

CHAPTER 3
Experiment 2[7]

3.1. Introduction

Supporting the generate-recognize approach (e.g., Bahrick, 1969, 1970; Anderson & Bower, 1972), Experiment 1 showed that participants indeed have a tendency to withhold some correct responses when they adopt a stringent report criterion. The experimental design herein was, however, changed not only to discover further but stronger evidence on the observation gained in Experiment 1 (reduced report criterion would yield more correct responses), but also to investigate some well-known facilitative effects on memory (i.e., categorical relationship between to-be-remembered items) might not be comparably facilitative on metamemory performance, such as on monitoring. Therefore, following the generate-recognize approach, a newly-designed manipulation in the study list structure was introduced, which is to be used for all experiments from Experiment 2 to Experiment 6. This amendment was the manipulation on the existence of categorical associations amongst the to-be-remembered (t-b-r) items in paired associate learning.

3.2. The aims and expectations

Considering the ideas of Higham and Tam (2005), it was thought that the generated candidates should be appropriate enough to be correct as well as distinctive enough to be differentiated (recognized) amongst possible candidates in order to reach efficient remembering. At an intuitive level, the cognitive system seems to have a parsimonious (or an economic) nature, and it might be expected to have a restricted search set to generate candidate items so as to render the retrieval performance efficient as well as expeditious. For instance, it seems to be just a futile endeavour to generate some candidate items that would be impossible to be correct. That is, the items generated should have higher probability to be the target. Hence, the candidates should be highly probable to be the sought-after information as well as they should be distinctive enough to be detected (recognized) as a target item amongst the alternatives.

[7] This study was presented in a poster format at Medicine, Health, and Life Sciences Postgraduate Conference on 3rd June 2008 in Southampton, England United Kingdom.

Besides aiming to find a higher proportion of correct responses in a forced-report condition compared to a free-report condition, the current experiment also aimed to investigate the effects of inter-target association not only on memory but also on metamemory performance of cued and uncued recall of paired-associates. Specifically, the current study aimed to investigate memory and metamemory performance of recall at the same time and in one retrieval task, namely recall. In this sense, Experiment 2 is different than the study of Guerin and Miller (2008) in the way that their study compares memory performance for the categorized and uncategorized lists (single-item study lists) between two retrieval tasks, recognition and recall. However, despite Experiment 2 study accepts the claims of Guerin and Miller (2008) on the distinctiveness issue, it compares recall (generation) and recognition performance in a 'single retrieval task', just like the studies of Higham and Tam (2005, 2006; see also Higham, 2002). Unlike the studies of Higham and Tam (2005, 2006), however, the relatedness (e.g., semantic/ categorical association) was only between the target material, hence, the cue-target pairs were consistently weak. In this sense, the relative merit of cue reinstatement on generation (retrieval) and recognition (monitoring) processes will also be in consideration.

In conclusion, Experiment 2 aimed to investigate: (a) whether forced report yields more correct responses than free report, and under which conditions these responses could be gathered, (b) the effects of organization amongst to-be-remembered items on both generation (retrieval) and recognition (monitoring) processes, which is subsumed under the recall process, and (c) whether cuing is always helpful on recognition as it could be on generation process, in contrast to the well-known facilitative effect of cues on retrieval.

Therefore, a unique manipulation was introduced in order to reach the objectives mentioned above. The essence of the manipulation based on the construction of study lists with categorical associations 'between targets only'. To be more specific, the target words in the current study were arranged in a way that they were expected to facilitate the generation process but deteriorate the recognition process (detection) at the time of retrieval. To reach this dissociative effect, the level of association amongst the target items was manipulated. As a result, whereas the targets having high inter-target association (high-ITA) were the exemplars of various pre-determined categories (thereby they have semantic association between each other as well), the low-ITA targets had no categorical or semantic association whatsoever between each other. Further, the cue words were arranged in a way that they did not have any semantic or categorical association between targets or with any other cues in the study lists (see section 3.3.3. for further details on the study lists). Hence, it was expected that the generation of candidates would be facilitated through a restricted search set via letting the cognitive system search particular categories to retrieve target items (in high-ITA target recall). Nevertheless, since many of the targets would be in these particular categories, the distinctiveness of the targets generated would be less compared to categorically unrelated targets (low-ITA lists), which as a result could jeopardize the recognition process particularly in the recall of categorically related targets.

Thus, the experiment used a 2(report type: free vs. forced) X 2(recall type: cued vs. uncued) X 2(inter-target association level: high vs. low) mixed design with inter-target association level manipulated within participants, and with report type as well as recall type manipulated between participants. The dependent variables were memory performance, namely memory quantity and memory accuracy (Koriat & Goldsmith, 1996c), and confidence level on the response given at the time of test, both of which let us investigate monitoring performance of the participants alongside memory performance. The metacognitive monitoring performance measured on the basis of type-2 signal detection approach was taken as the measurement of the recognition performance. On the basis of type-2 SDT modelling (see e.g., Figure 4), it was assumed that the greater the difference between mean scores of the two distributions (e.g., *d'* or AUC), one of which is of correct responses and the other one is of the incorrect responses generated, the better is the monitoring performance of the participants (the ability to discriminate between correct and incorrect responses). In other words, a better recognition performance of the correct responses is inferred.

Based on the aims and ideas outlined earlier, the hypotheses of the current study were:

(a) as a further but stronger converging evidence to the results gained in Experiment 1, the forced-report participants were expected to have higher memory performance than the free-report participants; (b) participants were expected to have higher retrieval performance in high-ITA target recall compared to low-ITA target recall. However, their monitoring performance was expected to be lower than (or at least comparable to) low-ITA target recall; therefore, (c), participants were expected to have higher memory quantity in high-ITA target recall compared to low-ITA target recall, whilst they were expected to yield comparable (even lower) memory accuracy as well as comparable –or lower- AUC scores in high and low-ITA target recalls.

3.3. *Method*

3.3.1. *Participants*

One-hundred-and-one graduate and postgraduate students (age: *M* = 20.51, *SD* = 3.53) of the University of Southampton, England whose first language was English participated in the experiment. Each participant was compensated for his/her time with course credits or £5 cash payment. Twenty-two of the participants (21.8 %) were male and 79 (78.2 %) of them were female. Each of the participants were randomly assigned to one of the four groups: free-report cued recall (*n* = 26; age: *M* = 21.12, *SD* = 2.37; 10 male and 16 female), free-report uncued recall (*n* = 26, age: *M* = 20.08, *SD* = 1.16; six male and 20 female), forced-report cued recall (*n* = 25; age: *M* = 21.28, *SD* = 6.36; four male and 21 female) and forced-report uncued recall (*n* = 24; age: *M* = 19.54, *SD* = 1.47; two male and 22 female).

3.3.2. *Materials*

3.3.3. *Study lists*

Two lists that composed of 40 word pairs in each (e.g., cue-target) were constructed on the basis of The University of South Florida Free Association Norms (Nelson, McEvoy, & Schreiber, 1998) and Extended and Updated Category Norms of Battig and Montague (1969) (Van Overschelde, Rawson, & Dunlosky, 2004); see Appendix B for the study lists. The cue words in the two lists were identical but the target words paired with these cues were different. The target words in one of the study list were decided to be from several mutually-exclusive categories (high-ITA list), and the target words in the other list were neither categorically or semantically associated to any other target (low-ITA list). Specifically, the high-ITA list composed of the target words being five exemplars of eight various categories (such as ‘STEEL, IRON, SILVER, COPPER, GOLD’ for the ‘metals’ category). On the other hand, the low-ITA list involved the cue-target pairs in which none of the target words were categorically related to each other (i.e., ‘DOOR, DOLLARS, ARMY, COFFEE’, etc). In both of the lists, cue and target associations (categorical and semantic) were negligible, even between any two words (cue and target) not paired. Inter-target association in terms of categorical relatedness existed only in the high-ITA lists and only between every five words drawn from a particular category. Since the cue words were *identical* in the two lists that had different ITA levels (high vs. low), the same cue word that was paired with an exemplar of a particular category in one list (high-ITA list) was paired with a different word in the other list, which did not have any categorical relationship with any other target words (low-ITA list); see Appendix B.

In order to counterbalance the categories used, the high-ITA list was randomly divided into two halves (i.e., high-ITA List-1 and high-ITA List-2) with the criterion that each list had to contain 20 word pairs but, at the same time, composing of the target words that were the exemplars of four different categories. As a result of this division, target words in the high-ITA List-1 and the high-ITA List- 2 were 5 exemplars of 4 mutually exclusive categories: *metals, four-footed animals, colours*, and *furniture item*s in the high-ITA List-1; and, *articles of clothing, alcoholic beverages, members of clergy*, and *fruits* in the high-ITA List-2. After the division of the high-ITA list, the low-ITA list was divided into two halves in such a way that the cue words between the first and between the second halves of each ITA level were identical. As a result, the same cue words existed in the List-1 of high-ITA and List-1 of low-ITA list (e.g., SCHOOL-STEEL, VICTIM-IRON vs. SCHOOL-DOOR, VICTIM-DOLLARS) and the same cue words existed in List-2 of high-ITA and the List-2 of low-ITA (e.g., CIRCUS-SHIRT, OCEAN–PANTS vs. CIRCUS-ARMY, OCEAN-COFFEE). Therefore, each cue word was paired with a target word having categorical association with other four targets in the same half list of high-ITA level (List-1 or List-2) and paired with a target having no categorical association with other targets in same half-list of low-ITA (List-1 or List-2).

In order to understand whether two half lists of high-ITA differ in terms of mean exemplar dominancy of the targets, the following t-test analyses were conducted. These analyses were based on the scores which were in fact the percentage of 642 participants who gave a particular word as an exemplar for a particular category asked in the study of Van Overschelde, Rawson, and Dunlosky (2004). It was found that high-ITA List-1 and high-ITA List-2 differed marginally in terms of mean dominancy (M = 70.3, SD = 21.1; M= 54.5, SD = 21.6, respectively), $t(19) = 2.240$, p = .07. That is, when they were freely asked to report exemplars for particular categories, 70% and 55% of 642 participants in the study of Nelson, McEvoy, and Schreiber, (1998) reported the exemplars that are the target words in high-ITA List-1 and high-ITA List-2, respectively. Because the presentation of half-lists was counterbalanced across participants, the mean dominancy differed marginally between high-ITA List-1 and high-ITA list-2 was ignored; see Appendix B for the comparisons of mean exemplar dominancy between categories matched in each half list of high-ITA.

Additionally, in order to avoid a confounding effect of semantic association, the lists were arranged in a way that none of the cues were semantically associated to any other cue, none of the cues were semantically associated to any other target and none of the targets were semantically associated to any other target, except five of the targets that were in the same category if the list was a high-ITA list.

3.3.4. *Procedure*

Participants were tested individually in a quiet and dimly lit research station located in the School of Psychology of the University of Southampton, England. The study involved two study-test cycles in which the study phases were computerized and the test phases were paper-pencil. Each participant was provided with a written informed consent form before the experiment started and they were taken to the study after reading and signing it. Because the participants had two study-test cycles, the presentation orders of the ITA level (high vs. low-ITA) and the half-list (List-1 and List-2) were counterbalanced. The cue words in List-1 of high-ITA and List-1 of low-ITA were exactly the same, as was the case in List-2 of high-ITA and List-2 of low-ITA. Therefore, the counterbalancing yielded four different versions of the presentation order: (1) high-ITA List-1, then low-ITA List-2, (2) high-ITA List-2, then low-ITA List-1, (3) low-ITA List-1, then high-ITA List-2, (4) low-ITA List-2, then high-ITA List-1. The participants were randomly assigned to one of these four versions of the study.

In the study phases, each participant was presented with a total of 20 word pairs on a Macintosh computer via Runtime Revolution software program. Each word pair appeared on the screen for 3 seconds with 1-second inter-stimulus interval. The pairs were presented in the middle of the computer screen successively and in capital letters separated with a hyphen (i.e., TENNIS - SILVER). The presentation order of the word pairs was totally randomized by the computer program. Before the appearance of the words on the screen, the participants were instructed to attend to all of the words in the to-be-presented list because they would be required to try and remember as many words as possible after the presentation of the list was complete.

In the test phases, each participant was given two-paged reporting sheets to be used to report the target words they remembered (see Appendix B). On the first page, there were instructions specific to the condition to which the participant was allocated. On the second page (the reporting page), there were two columns: ‘CUES’ (written down one under the other) and ‘TARGETS’. The ‘TARGETS’ column contained empty spaces adjacent to the 'CUES' column if the participant was in a cued-recall group. On the other hand, if the participant was allocated to an uncued-recall group, there was not a 'CUES' column, but only a 'TARGETS' column. Whereas the free-report participants were free to report and withhold as many words as they wished, the forced-report participants were asked to complete all of the 20 empty spaces of the 'TARGETS' column. The participants also defined their confidence levels for each word they reported to indicate how confident they were that the response was correct (target). They rated their confidence levels on a Likert-type scale given next to each response and ranging between 1 and 7 (1 = "Not at all confident correct", 4 = "Fairly confident correct", 7 = "Completely confident correct").

None of the participants knew whether they would be provided with cues or not in advance of the study phases. They were also not informed in advance about which side of the pairs (cues or targets) needed to be recalled. Hence, they learned which words to be recalled (targets) and which ones would be used as cues (valid only for cued-recall groups) at test phases. Whereas the testing method both for the two study lists depended on the group to which the participant was allocated (cued vs. uncued recall, and free vs. forced report), the way of studying the pairs was identical for all participants. Although the participants were either asked to remember the targets with cues or without cues at testing, they were all asked to remember and report 'only the target words' which were the words on the right side in the pairs studied. Each participant was warned in the instructions preceding the presentation of the second list that they did not know how the presentation of the second list words would be to eliminate the effect of the knowledge which could be gained by the first testing on the nature of the testing method (i.e., which words would be reported, whether they would be provided with cues or not). Therefore, they were again asked to fully attend to the word pairs and were informed that they would be responsible to remember as many words as possible for the second list as well.

The experiment ended when the second testing phase was completed. After the second testing phase, they were given a written debriefing statement and their possible enquiries were responded. Since each testing phase was self-paced, the completion time of the study ranged between 25-45 minutes.

3.4. Results

3.4.1. *Memory quantity*

The scoring of responses was either strict or liberal. The *strict scoring* referred to counting ‘only the targets paired with their studied cues’ as correct responses. On the other hand, the *liberal scoring* referred to the target words being counted as correct even if they were not matched with the cues paired at study. Therefore, strict and liberal scorings were valid for only the cued-recall participants. From Experiment 2 to Experiment 6, liberal scoring was adopted for all calculations. All post-hoc mean comparisons in Experiment 2 were Bonferroni corrected. Table 7 displays the mean confidence levels given for targets, cues, and extra-list words reported as well as the number of intrusion errors (cues and extra-list) as a function of inter-target association in each group.

First, a 2(report type: free vs. forced) X 2(recall type: cued vs. uncued) X 2(ITA level: high vs. low) mixed ANOVA was conducted on memory quantity (MQ)[8]; see Table 6. The results revealed a report type main effect, $F(1, 97) = 11.631$, $p = .001$, $\eta 2 = .107$. The forced-report participants had significantly higher correct responses (M = 59.1, SE = 2.1) than the free-report participants (M = 48.6, SE = 2.2). This result confirmed the hypothesis: Reducing recall criterion had a significant increase on correct responses reported. In other words, participants withhold some correct responses that they had due to stringent report criterion they adopted under free report. Recall type main effect was also significant, $F(1, 97) = 37.644$, $p < .001$, $\eta^2 = .280$: The cued-recall participants reported significantly more correct responses (M = 63.3, SE = 2.2) than the uncued-recall participants (M = 44.4, SE = 2.2). This result converged to the encoding specificity principle (Tulving & Thomson, 1970) that basically proposes the cues are more effective when they are available both at the time of study and at the time of test. Report type and recall type did not have an interaction effect on MQ, $F(1, 97) = 1.076$, $p > .05$. ITA main effect was significant, $F(1, 97) = 10.317$, $p = .002$, $\eta 2 = .096$; participants had greater memory performance in high-ITA target recall compared to low-ITA target recall (M = 56.9, SE = 1.7; M = 50.8, SE = 1.9, respectively). The results also showed that there was an ITA level and recall type interaction effect on MQ, $F(1, 97) = 9.245$, $p = .003$, $\eta^2 = .087$. Pair-wise mean comparisons revealed the

[8] The results gathered when MQ was measured with strict scoring revealed exactly the same pattern gathered on memory quantity measured with strict scoring (MQ-str), except ITA level main effect was not significant, $F<1$.

cued-recall participants did not differ in terms of MQ between high-ITA target recall (*M* = 63.4, *SD* = 21.4) and low-ITA target recall (*M* = 63.1, *SD* = 23.4), $t(50) = .097$, $p > .025$; however, the uncued-recall participants had higher MQ in high-ITA target recall (*M* = 50.3, *SD* = 14.0) than low-ITA target recall (*M* = 38.6, *SD* = 13.9), $t(49) = 5.488$, $p < .001$. This finding is important to show the following implication: If the cues are available at retrieval, the existence of categorical association between targets does not turn out to be an essential factor to have a facilitative effect on recall performance. On the other hand, if the cues are not overly available at testing, the categorical relatedness amongst targets facilitates recall. Lastly, there was no interaction effect between ITA level and report type and between ITA level, report and recall types, both *F*s<1.

3.4.2. *Memory accuracy*

A 2(report type: free vs. forced) X 2(recall type: cued vs. uncued) X 2(ITA level: high vs. low) mixed ANOVA was conducted on memory accuracy; see Table 6. The results revealed a report mail effect, $F(1, 97) = 70.430$, $p < .001$, $\eta^2 = .421$. The free-report participants were more accurate in their responses (*M* = 83.6, *SE* = 2.0) than the forced-report participants (*M* = 59.1, *SE* = 2.1). This was inevitably expected since the forced-report participants were asked to report same number of responses that they studied (20) even if they needed to make guess, unlike the free-report participants. The results also revealed a recall type main effect, $F(1, 97) = 12.424$, $p = .001$, $\eta^2 = .114$: The cued-recall participants had greater MA than the uncued-recall participants (*M* = 76.5, *SE* = 2.1; *M* = 66.2, *SE* = 2.1, respectively). There was not an interaction effect between report and recall types, $F(1, 97) = 3.332$, $p > .05$. However, ITA main effect was significant, $F(1, 97) = 12.407$, $p = .001$, $\eta^2 = .113$; Participants were more accurate in high-ITA target recall compared to low-ITA target recall (*M* = 74.8, *SE* = 1.6; *M* = 67.9, *SE* = 1.9). Whilst there was not a significant interaction effect between ITA and report type, *F*<1, the interaction effect between ITA and recall type was significant, $F(1, 97) = 5.542$, $p = .021$, $\eta^2 = .054$. The cued-recall participants did not differ in terms of accuracy between high-ITA and low-ITA target recalls (*M* = 77.5, *SD* = 21.0; *M* = 75.4, *SD* = 22.8, respectively), $t(50) = .891$, $p > .025$. On the other hand, the uncued-recall participants were more accurate in high-ITA target recall (*M* = 72.0, *SD* = 19.5) than in low-ITA target recall (*M* = 60.4, *SD* = 22.9), $t(49) = 3.963$, $p < .001$. Lastly, there was not a significant interaction effect between ITA level, report type and recall type on memory accuracy, $F<1$.[9]

Table 6

Means and Standard Deviations of Memory Quantity and Memory Accuracy as a Function of Inter-Target Association Level, Report Type, and Recall Type in Experiment 2

		Memory quantity		Memory accuracy	
		Inter-target association level			
Report type	Recall type	High	Low	High	Low
Free	Cued recall	59 (22)	60 (25)	88 (17)	85 (20)
	Uncued recall	43 (12)	32 (12)	86 (14)	76 (19)
Forced	Cued recall	68 (20)	66 (22)	-	-
	Uncued recall	58 (12)	45 (14)	-	-

Note. Standard deviations are in parentheses. Memory quantity and memory accuracy performance were exactly same for the forced-report groups. Only for displaying purpose, these performance were shown under memory quantity for the force-report groups.

Table 7

Mean Confidence Levels (CF) Targets, Cues, and Extra-list Words and the Number (Nbr) of Cue and Extra-List Intrusion Errors Reported As a Function of Semantic Association Level (High vs. Low), Recall Type (Cued vs. Uncued)

			Target		Cue		Extra-list	
Recall	Report		High	Low	High	Low	High	Low
Cued	Free	CF	6.3 (0.8)	6.3 (0.8)	1.3 (0.4)	2.7 (1.4)	2.1 (1.3)	2.3 (1.4)
		Nbr			0.0 (0.2)	0.3 (0.9)	1.7 (2.7)	1.8 (2.5)
	Forced	CF	5.9 (1.0)	6.1 (1.2)	1.3 (0.5)	1.6 (0.8)	1.9 (1.8)	1.3 (0.9)
		Nbr			0.6 (0.9)	0.6 (1.0)	5.9 (3.9)	6.0 (4.0)
Uncued	Free	CF	5.7 (1.0)	5.7 (1.1)	4.6 (0.9)	3.8 (1.6)	3.6 (1.8)	2.9 (1.4)
		Nbr			1.2 (1.6)	1.7 (1.4)	0.3 (0.5)	0.3 (0.7)
	Forced	CF	4.9 (1.2)	4.4 (1.3)	2.1 (1.0)	3.0 (1.3)	1.3 (0.4)	1.3 (0.4)
		Nbr			4.2 (1.7)	5.1 (2.5)	4.3 (2.2)	5.9 (4.0)

Note. Standard deviations are in parentheses. See Table 6 for the percentage of targets reported.

[9] The results obtained when strictly scored MA was analysed revealed the same results as the results when MA-lib was analysed, except recall type as well as ITA level did not show any main effects, both Fs<1.

3.4.3. *Metamemory performance: monitoring*

Monitoring performance, measured on the basis of type-2 signal detection theory (cf., Higham, 2002), was explored in the following analyses. The same calculation method of area under the curve in Experiment 1 was utilised here as well (see section 2.13. for details of AUC calculation). Figure 6 displays the ROC curves constructed on the basis of type-2 hit and false alarm rates.

A 2(report type: free vs. forced) X 2(recall type: cued vs. uncued) X 2(ITA level: high vs. low) mixed ANOVA on AUC scores was conducted; see Table 8. The results indicated report type main effect was not significant, $F(1, 97) = 1.998$, $p > .05$. However, recall type indicated a main effect, $F(1,97) = 13.038$, $p < .001$, $\eta^2 = .118$: The cued-recall participants monitored their responses better than the uncued-recall participants ($M = .91$, $SE = .02$; $M = .82$, $SE = .02$, respectively). There was not an interaction effect between report and recall types, $F<1$. The results did not also reveal a main effect of ITA level. There was not a significant interaction effect between ITA and report type, $F<1$, the interaction effect between ITA and recall type was, however, significant, $F(1, 96) = 4.090$, $p = .046$, $\eta^2 = .040$. Post-hoc mean comparisons showed that the cued-recall participants monitored their responses better in low-ITA target recall compared to high-ITA target recall ($M = .93$, $SD = .09$; $M = .88$, $SD = .18$, respectively) at a marginally significant p value, $t(50) = 1.924$, $p = .06$. On the other hand, the uncued-recall participants monitored their responses comparably well in low-ITA and in high-ITA target recalls ($M = .83$, $SD = .16$; $M = .81$, $SD = .19$, respectively), $t(49) = .785$, $p > .05$. Lastly, there was a three-way interaction effect was between ITA level, report type, and recall type on monitoring performance, $F(1,97) = 6.572$, $p = .012$, $\eta^2 = .063$. Post-hoc mean comparisons showed that whereas free-report cued-recall group did not differ in terms of monitoring performance between high-ITA and low-ITA target recalls ($M = .92$, $SD = .11$; $M = .93$, $SD = .09$, respectively), $t(25) = -.247$, $p > .01$, just like free report uncued-recall participants ($M = .82$, $SD = .18$; $M = .85$, $SD = .20$, respectively), $t(25) = -.597$, $p > .01$, and forced-report cued-recall participants, ($M = .83$, $SD = .22$; $M = .93$, $SD = .09$, respectively), $t(24) = -.2.183$, $p > .01$, forced report uncued-recall participants monitored their responses better in high-ITA target recall ($M = 85$, $SD = .12$) than in low-ITA target recall ($M = .77$, $SD = .16$), $t(23) = 2.841$, $p < .01$.[10]

[10] The results gathered when strictly scored AUC scores were analysed showed exactly the same pattern when liberally scored AUC scores were analysed, except ITA and recall type did not show an interaction effect, $F(1, 97) = 3.063$, $p > .05$.

Table 8

Means and Standard Deviations of Area Under The Curve Scores as a Function of Inter-Target Association Level, Report Type, and Recall Type in Experiment 2

		Inter-target association level	
Report type	Recall type	High	Low
Free	Cued recall	.92 (.11)	.93 (.09)
	Uncued recall	.82 (.18)	.85 (.20)
Forced	Cued recall	.83 (.22)	.93 (.09)
	Uncued recal	.85 (.12)	.77 (.16)

Note. Standard deviations are in parentheses.

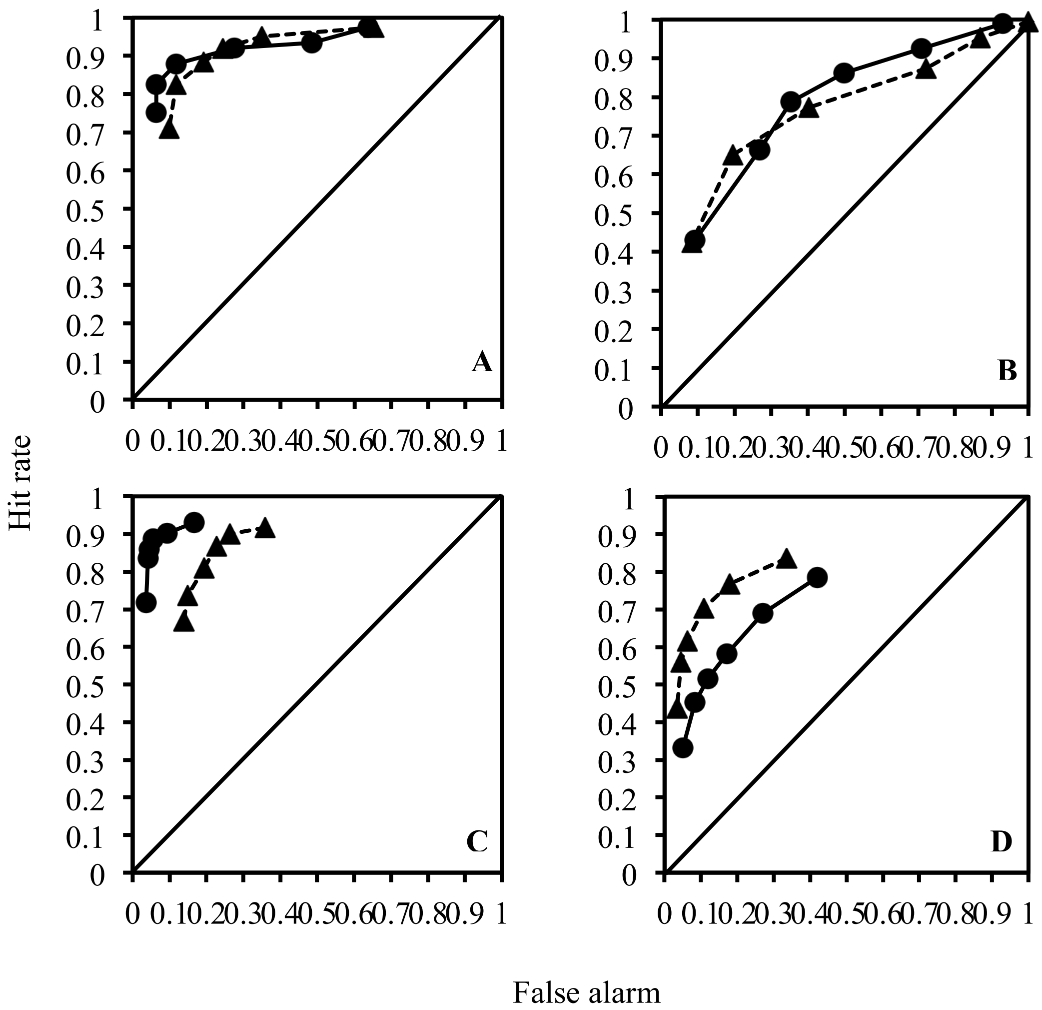


Figure 6. ROC curves displaying the intersection points of hit and false alarm rates at cumulative confidence levels of high-ITA and low-ITA target recalls in the free-report cued-recall (A), the free-report uncued-recall (B), the forced-recall cued-recall (C), and the forced-report uncued-recall groups (D). Straight lines (—) display ROCs in high-ITA target recall, whilst dashed lines (····) display ROCs in low-ITA target recall.

3.5. Discussion

The results revealed that reducing recall criterion increases overall correct responses. As found in Experiment 1, this main finding disconfirms the conclusions of Roediger and Payne (1985) via empirically supporting that the basic assumption of generate-recognize theories is not false and people seem to have more information in free report than they produce. In other words, participants had possibly more correct responses than they reported, however, they seemed to refrain from reporting some of them -if not all- because of a stricter report criterion they had when they were free to report. Converging to this conclusion again, those who were forced to report, even though they were required to give their best guesses, had higher percentage of correct responses compared to the free-report participants who reported less for the sake of being more accurate in their responses.

Why could the current study confirm that reduced recall criterion has an incremental effect on overall correct responses, but the others could not? The reason behind this difference could primarily be attributed to the different methodology, combined with the learning paradigm utilised in the previous studies that compared free and forced report performance.

First, let me consider the two broad categories of testing methodology comparing free-report and forced-report performance. For instance, in some studies free-report and forced-report options were compared *between participants* (i.e., Dywan & Bowers, 1983; Roediger & Payne, 1985). On the other hand, some studies compared these performances *within participants*. In one sub-category of the within-participant comparisons, for instance, participants were first asked to free report, then asked to indicate the last response they gave (such as by drawing a line under the last word they free-reported), and then asked to go on reporting even if they need to make pure guesses (e.g., see Erdelyi, Finks, & Feigin-Pfau, 1989). In some other studies with a within-participant comparison, however, participants were allowed to decide as to whether they want to report their responses or pass to give a response for every cue, item, space, or question provided (see, e.g., Higham, 2002; see also, Higham & Tam 2005, 2006; Higham & Arnold, 2007). In the latest methodology, the responses are given either with a 'report' option ("I am confident that I remember the response") or with a 'pass' option ("I am not confident to give a response/answer since I do not remember/know anything, but I have to make a guess"). Based on this report-or-pass methodology, the number of responses given with 'report' option constitutes the free-report performance, and the total number of responses given with 'report' as well as 'pass' options constitutes the forced-report performance.

As Erdelyi, Finks, and Feigin-Pfau (1989) suggested, comparing free-report and forced-report options between participants could lead to the 'processing bias', termed by Erdelyi (1985). That is, participants might divide their effort across all items in forced reporting. However, when participants are tested within participants in the method of 'free report-draw a line-continue', participants might be expected to have dumped all of the information they could have in free report part of the procedure. If participants already report (or dump) all the information they have under free-report option, then it seems unreasonable to expect further correct responses being reported under a following forced-report condition (cf. Cofer, 1967). However, it is quite reasonable to expect that not all information is reported under the free-report option, since some of the responses (or information) would expectedly vary in terms of confidence levels. Although it might be an issue whether there is a perfect match between subjective confidence levels attached to the confidence (or strength) of correct and incorrect responses (cf., Higham, Luna, & Bloomfield, 2011), it has almost always been observed that participants do not either report their responses as all attached with the highest confidence level available or with the lowest confidence level available in the confidence rating scales (e.g., see ROC curves in Figure 6). In short, the testing methodology in this experiment, which compared the free-report and the forced-report performance between participants, seems to be a factor to yield a difference, although it does seem that it has not created the difference alone. In other words, the reason for the difference found between participants in terms of gathering more correct responses in forced-report conditions (particularly in uncued recall), unlike earlier experiments (e.g., Dywan & Bowers, 1983; Roediger & Payne, 1985), is more understandable when the learning paradigm and the nature of the study material utilised in the current experiment is considered.

The present study used paired-associate learning paradigm unlike the previous studies, which used single-item study lists heavily (e.g., Cofer, 1967; Dywan & Bowers, 1983; Roediger & Payne, 1985, 1989; Erdelyi, Finks, & Feigin-Pfau, 1989). For instance, when two conditions in which study lists involve the same number of to-be-remembered materials are considered, participants might be expected to spend more effort in paired-associate learning tests (in which study lists compose of cue and target pairs) both at the time of study to encode the material and/or at the time of testing - where the latter might depend on whether the cues are provided to participants or not-compared to those ones using single-item study lists (in which study lists compose of single to-be-remembered items). Therefore, finding a facilitative effect of reduced recall criterion (e.g., under forced report) on the number of correct responses reported seems to be an important factor when combined with the nature of learning paradigm; paired-associate. However, this paradigm, as a factor on gathering more correct responses when participants were forced to report, also combines with the testing procedure; whether participants are cued or not. That is, providing participants with cues seems to be helpful in remembering the target words. The results of Experiment 2 showed that *when participants were provided with cues,* free and forced-report retrieval performance of targets did not differ. That is, the participants seemed to simply rely on the availability of cues as a strong base to remember the to-be-remembered items. However, a significant difference was found only when *uncued-recall participants* were analysed: The forced-report uncued-recall participants had significantly higher memory quantity than the free-report uncued-recall participants. Why was such a result gathered only between free and forced report participants who were *uncued* at the time of testing?

It is necessary to keep in mind the fact that the uncued-recall participants either in free-report or forced-report groups had no (overt) help of cues at the time of testing, and the only difference between them was the criteria to report. Therefore, as participants studied the pairs as 'pairs' because of the fact that none of the participants were informed about the places of the targets (which items to be remembered), the first possibility could be based on the fact that whichever group remembered more cues might have remembered the targets paired with them much more easily. The results showed that the forced-report uncued-recall participants reported a significantly greater number of cues (M = 23.13, SD = 7.85) compared to the free-report uncued-recall participants (M = 7.31, SD = 5.24); $t(48) = 8.444$, $p < .001$. However, this result might not directly reveal the reason as to why the forced-report participants remembered more target words compared to the free-report uncued-recall participants. It was, however, speculated that reporting more cues might have facilitated the retrieval of the targets paired with them amongst the forced-report uncued-recall participants.

Besides recall type (cued vs. uncued), the facilitative effect of reduced recall criterion on yielding more correct responses was further investigated by taking ITA level into consideration. The results showed that the forced-report uncued-recall participants had significantly higher correct responses than the free-report uncued-recall participants, regardless of whether the targets were categorically related to each other or not. The difference gathered in high inter-target association seems understandable because the forced-report participants could have taken advantage of the target association, either based on a guidance to generate more exemplars from particular categories studied (which could possibly be a target word), or a target word remembered might have activated (or eased) the retrieval of another target word being in the same category. Alternatively and most possibly, this difference might be the case depending on both of the reasons. Put simply, any categorically-related words are semantically related to each other. However, not all semantically-related words are categorically related. For instance, '*bed*' and '*dream* ' are related to each other in terms of semantic association, however, '*bed*' is both semantically related to '*table*' and in the same category of such as '*articles of furniture*'. The idea here is based on the construction of 'close relationships*'*, otherwise the following possibility always exists: Anything might be linked with anything else in terms of a categorical relatedness (e.g., '*bed*' and '*dream*' might be considered in the same category, such as '*sleeping*' etc.). Therefore, any target word retrieved on the basis of a guidance which could restrict the search set might have also activated the other target words semantically-associated with this target. As a result, any retrieved target word might have eased the retrieval of other targets in the particular category searched.

However, the higher memory quantity in low-ITA target recall gathered in the forced-report uncued-recall participants compared to the free-report uncued-recall participants does not seem to be explained by the reason(s) explaining high-ITA target recall. Because, the low-ITA targets did *not* have neither categorical nor semantic associations between each other so as to facilitate the retrieval of other targets. Hence, low-ITA targets lacked the advantage that participants had in high-ITA target recall such as, the guidance in searching as well as the *spread of activation* (e.g., cf. Anderson & Bower, 1973) which might have eased the retrieval of other related targets. The reason for this finding might simply be come about from the difference in the participant characteristics between the free-report and the forced-report uncued-recall groups. However, another possibility could be as follows. The *free-report* uncued-recall participants might have set a 'quite stringent report' criterion in low-ITA target recall, both due to not having cues at the time of retrieval, which is also disadvantageous in terms of not having the target association, as well as not knowing the place of the cues and targets in advance. Put it differently, these conditions together –if not separately- could have allowed these participants to set more stringent report criteria compared to when the testing was for high-ITA targets.

In short, the reasons for gathering more correct responses in high-ITA target recall and in low-ITA target recall in the forced-report uncued-recall participant compared to the free-report uncued-recall participants were respectively attributed to the existence of inter-target association; and, possibly more stringent report criterion set in low-ITA target recall compared to high-ITA target recall amongst the free-report uncued-recall participants. That is, gathering of higher correct responses in low-ITA target retrieval amongst the forced-report uncued-recall participants might be illusory in the way that the free-report uncued-recall participants might have failed to report all of the responses, due to a quite stringent report criterion.

In conclusion, the reason behind finding higher correct responses amongst the forced-report compared to the free-report participants seem to be a combined factor of the learning paradigm (recall type: cued or uncued), as well as inter-target association level.

More importantly, however, besides disconfirming the conclusions that seem to undermine the assumptions of the generate-recognize approach, Experiment 2 found more evidence on the theory's assumptions, and confirmed them once more. That is, the results of Experiment 2 showed that the inter-target association facilitated retrieval performance particularly in uncued recall both under the free-report and the forced-report groups. However, as was expected, the facilitative effect of target association on retrieval did not reflect on monitoring performance (e.g., AUC scores) in the same way. Therefore, it seemed that the organized lists facilitate the generation process but they do not have the same facilitative effect on monitoring. In this sense, there existed a partial dissociation between memory and metamemory performance as a function of ITA level in Experiment 2. However, a full dissociation between memory and metamory performance (e.g., a facilitative effect of inter-target association on memory along with a subsequent deteriorating effect on monitoring performance) can be expected as the number of categories in high-ITA list is reduced (see Experiment 5 which manipulates the number of categories, e.g., two vs. six vs. twenty-four, and finds a full dissociation between memory and metamemory performance due to utilization of high inter-target association).

In summary, the results of the Experiment 2 showed that the type of testing (whether free or forced), the structure of the study lists which were created to have categorical relationship amongst target items or not, and the availability of the cues at test or not (cued vs. uncued recall) together with studying the cue-target pairs without knowing the places of them in advance work together - if not alone- as a facilitative or as a deteriorating factor not only on memory, but also on metamemory performance. Amongst these findings, the most central result of the study was that organization, endorsed as a well-documented facilitative effect on memory, does not seem to have the same merit on metacognitive monitoring.

CHAPTER 4
Experiment 3

4.1. The aims and expectations

Experiment 2 confirmed that participants report more correct responses when they were forced to report compared to when they were free to report, particularly when they were not provided with any cues at the time of testing (uncued recall). This result was important to support the predictions of generate-recognize approach by pointing out that participants withhold some correct responses that they generated because of a stringent report criterion employed in free report. Observing higher numbers of correct responses in forced-report compared to free-report particularly in uncued recall also seemed to converge with the findings of organizational effects becoming salient in free (uncued) recall rather than in cued recall. That is, although a few studies found organizational effects in paired-associate learning (e.g., Battig, 1966; Runquist, 1970), these effects have traditionally been investigated and clearly observed in free-recall experiments (e.g., Tulving, 1962; Tulving & Pearlstone, 1966; Puff, Murphy & Ferrara, 1971). Gathering greater retrieval performance when targets were inter-related in forced report compared to free report might naturally be expected since such inter-relatedness might have led uncued-recall participants to use category belongingness amongst the to-be-remembered (t-b-r) items. It should, however, be noted that Experiment 2 also showed that a forced-report option could yield higher correct responses than a free-report option in uncued recall even though target items did not have any experimentally designated inter-relatedness (e.g., low-ITA target list). This result was attributed to the possible highly stringent report criterion adopted for low-ITA target recall under free-report testing. More importantly, however, Experiment 2 demonstrated that having categorical relationships between targets resulted in a facilitated retrieval of these targets, but this facilitation did not appear in a parallel way for monitoring performance. This dissociation –although partial- supported the expectation of the current research: The generation processes could be facilitated with the categorical relatedness between targets, however, it did not enhance metacognitive monitoring in a parallel way.

Following Experiment 2, Experiment 3 aimed to understand under which conditions inter-target association is utilised best to facilitative retrieval of related targets and thereby monitoring performance would be expected to be affected negatively or, at least, not parallel to the facilitation in memory performance. It was thought that study and retrieval contexts as well as the combinations of them would lead the utilization of ITA in retrieval that thereby could result in a predicted facilitation in memory performance along with monitoring performance that is not enhanced in the same way. The following paragraphs explain the critical manipulations to reach the expected dissociation between memory and metamemory performance due to the utilization of inter-target association.

The *first* manipulation in the current experiment was made on the encoding task (study context). At this point, the levels of processing idea of Craik and Lockart (1972), seems well-suited to understand enhancements in retrieval, and mainly predicts that the deeper the information is processed, the better the retrieval performance. One class of empirical supports for the levels of processing idea of Craik and Lockhart (1972) comes from the memory research using the *incidental-learning paradigm*. The paradigm conventionally refers to the testing of memory in which participants study the given material via experimentally designated orienting tasks and characteristically without expecting to learn the material so as to be tested on it in the future (for a review see e.g., McLaughlin, 1965). In this sense, incidental-learning experiments could be contrasted with the *intentional-learning* experiments in which participants are informed in advance that they would be tested on the learned material in a prospective remembering task. The orienting tasks used in the incidental-learning studies, however, referred to tasks that are typically defined by the experimenter with an intention to manipulate the levels of processing participants use whilst studying the material. In other words, these tasks are assumed to vary the processing of the information which necessitates either a shallow processing or a deeper processing, such as judging the structural features of the study words or deciding whether the study words have a categorical relationship with a concept given by the experimenter (Tresselt & Mayznet, 1960; see also, Craik & Lockhart, 1972; Craik & Tulving, 1975; Shulman, 1971; Koriat & Melkman, 1987). Although having some contradictory findings (Slamecka & Gnaf, 1978), the levels of processing approach has been supported by some investigations regarding recall (e.g., Tresselt & Mayznet, 1960; Craik & Tulving, 1975; Mantyla, 1986; Mantyla & Nilsson, 1988; Hunt & Smith, 1996), recognition performance (e.g., Stein, 1978; Shulman, 1971), memory organization (Koriat & Melkman, 1987; see also Hunt & Smith, 1996 for the effectiveness of cues in recall with respect to distinctive processing of categorized lists) as well as in serial learning paradigm (Polyn, Norman, & Kahana, 2009a, 2009b). In short, the findings in general are in favour of the levels of processing perspective.

Unlike incidental learning experiments, in the experiments using the paired-associate learning paradigm that are based on intentional learning, participants are conventionally informed which items are to be remembered (targets) and which ones would be utilised to aid the retrieval of the target (cues). In the current experiment, however, the type of studying (encoding) the pairs were expected to pay a role by letting participants either depend on any inter-target association exist in the study lists or not. As a result, the expectations regarding the possible dissociative effect of ITA on memory and metamemory were expected to depend on whether these associations would be utilised at retrieval or not.

The encoding strategy in the current experiment was, therefore, manipulated in terms of the types of study instruction which aimed to vary the focus of the participants whilst studying the paired items: pair-focused vs. target-focused. The *pair-focused group* was not informed about the places of the words in advance, such as the cues are the words on the left hand side in the pairs, whereas to-be-remembered words are on the right hand side. Instead, they were asked to study the pairs as they would be responsible for remembering as many words as possible from the just-presented list. On the other hand, the *target-focused group* studied the word lists knowing that they would be responsible for remembering the target words presented on the right hand side in the pairs, by using the cue word to aid retrieval in a future recall test. As the manipulation implies, Experiment 3 is based on the intentional-learning paradigm since both groups were aware (informed in advance) that they would be tested in a subsequent recall test. However, the only manipulation was on the participants' focus during studying the pairs. More importantly, besides being the first study directly comparing the effects of different study type instructions (e.g., pair-focused vs. target-focused ways) in cued as well as free (uncued) recall of paired-associate material, the present experiment aimed to investigate the effects of inter-related study materials on memory and metamemory performance with regards to study type.

The expectations for the cued-recall and the uncued-recall group in terms of the organizational effects (e.g., categorization) on memory and metamemory performance were different from each other. In uncued recall, it was expected that studying the lists with either paired-focused or target-focused way would both result in higher memory quantity for high-ITA target recall compared to low-ITA target recall, although target-focused studying would reveal even higher memory quantity both for high- and low-

ITA target recalls as opposed to pair-focused. This prediction was based on the idea that when participants are not provided with any 'covert cue' at testing (uncued recall), they would be seeking for available sources at retrieval, such as inter-target association. In cued recall, however, it was expected that lowered monitoring performance would be gathered depending on whether the existed inter-target relationship is expected to be utilised or not. In other words, it would depend on whether study type would let participants form organizations such as higher-order units (i.e. realizing category relation between targets) or not. Therefore, the target-focused cued-recall participants were predicted to be able to better realise the categorical relationship between targets, which in turn would facilitate the formation of conceptual (even if not only semantic) organization compared to the pair-focused cued-recall group.

Koriat and Melkman (1987), who investigated various levels of processing in memory organization, found that conceptual organization emerged when participants studied the lists of words with a deeper encoding. That is, organizing the material in terms of taxonomic characteristic necessitates more effort compared to organizing them on the basis of their semantic features. Based on this finding, it was thought that the probability of realizing the categorical relationship between targets that exists in the study list would be higher in the target-focused group compared to the pair-focused group. Being asked to study the word pairs as 'pairs' would presumably necessitate more effort to encode the material as pairs compared to when the participants know that they would be responsible only for the targets (the right hand side words in the pairs). The differentiated level of processing between study types seems more understandable when the following is considered: the study time for each word pair was constant (3 sec) across all conditions. Therefore, when equal study time allocated to each pair both in target-focused and pair-focused way of encodings is considered, the encoding of preferably single words (targets) compared to two words (cue-target pairs), was expected to yield deeper encoding of the targets along with a less connected encoding with the cues given. It was, therefore, hypothesized that (expectedly) facilitated recall performance in categorically related targets along with a not facilitated or even lowered monitoring performance of these inter-related targets would be observed in the target-focused cued-recall group rather than in the pair-focused cued-recall group. Because conceptual organization that could be observed in terms of retrieving a higher number of categorically related targets through better realizing the relationship between them is expected amongst the target-focused participants rather than in the pair-focused participants. In parallel with this expectation, the pair-focused cued-recall group was predicted to take advantage of cue existence to a higher extent than the target-focused cued-recall group. This prediction would be observed/measured with an expectedly lower percentage of mismatched targets in the pair-focused cued-recall group compared to the target-focused cued-recall group.

The *second* unique manipulation in the present study was at the time of testing (retrieval context). The cued-recall participants were either instructed that their responses would be counted correct only when they make the correct match with the cues and responses given, just as they were paired at study. This group of participants constituted the *constrained-cued-recall* participants. On the other hand, the *liberated-cued-recall* group were instructed that their responses would be counted correct regardless of the fact that whether the target response they give for the cue provided was the same cue paired with at study or not[11].

The scoring manipulation at retrieval was introduced since it served three purposes. First, it has been well documented by the encoding specificity principle (Thomson & Tulving, 1973) and *transfer appropriate processing* (Morris, Bransford & Franks, 1977) that matching study and retrieval contexts yields better memory performance. Therefore, it was thought that whereas constrained retrieval is more compatible for the pair-focused cued-recall group, liberated retrieval context is a more compatible retrieval context for the target-focused cued-recall participants. In other words, when participants focus on targets more during studying the pairs, testing their retrieval performance via instructing that they could ignore which cue the response was paired at study (liberated-cued recall) was more congruous compared to a condition that they had to consider matching the right response (target) with the very cue that was paired with at study (constrained-cued recall). It was, therefore, expected that the pair-focused cued-recall group would yield comparable memory performance between high and low ITA target recalls when they were either constrained or liberated in terms of cue-target matching. In other words, they were expected to refrain from liberation and would behave as those in the constrained-cued recall since a pair-focused encoding strategy would let them depend on cues more in retrieval as the encoding specificity principle suggests. On the other hand, the target-focused cued-recall group was expected to yield higher retrieval performance for categorically related targets compared to categorically unrelated ones when they were liberated in terms of cue-

[11] Since only the cued-recall participants are provided with the cue words at the time of testing, the scoring instructions pertaining to cue-target matching (constrained vs. liberated) were only valid for the cued-recall participants.

target matching compared to when they were constrained in doing so. Second, if the target-focused cued-recall group yield comparable retrieval performance (e.g., forced-report MQ) in constrained and liberated retrieval conditions, then the results would be attributed to the fact that the target-focused group depends on cue availability at retrieval rather than inter-target association. Third, as it was observed in Experiment 2, the percentages of mismatched targets given at testing increased when targets were categorically related. Hence, a specific prediction to observe the clearest dissociative pattern in cued recall could be made. That is, the highest probability of observing the inter-target association effect on retrieval and on monitoring in cued-recall group was expected in the liberated-cued recall group who studied the pairs in a target-focused way.

In short, when study and test contexts are considered together, the predictions for the cued-recall participants varied as a function of the interaction (combinations) between study type (pair-focused vs. target-focused), scoring instruction (constrained vs. liberated). Table 9 displays the specific predictions in the cued-recall group as a function of whether the generation processed is expected to be higher (observed via forced-report MQ) in high-ITA target recall compared to low-ITA target recall so that a subsequent deteriorated –or at least, comparable- monitoring performance could be expected. Based on Table 9, a possible dissociation between memory and monitoring is expected only when a facilitated generation is observed.

Table 9

*Whether Categorically Related Targets are Predicted to have Higher Retrieval Compared to Categorically Unrelated Targets (*Yes-No*) as a Function of Study Type and Scoring Instruction in Cued-Recall*

	Scoring instruction	
Study type	Constrained cued recall	Liberated cued recall
Pair-focused	No	No
Target-focused	Yes	Yes

In addition to the scoring instructions, the cue words given at the time of testing were arranged in a way that when the targets are perfectly matched with the same cues they are paired with at study, these targets belonging to the same category come one after another. That is, they create blocks of exemplars subsumed under particular categories. It was thought that such an arrangement would boost the retrieval of the category exemplars. To be more specific, the following was expected. Any retrieved target(s) would ease the retrieval of another related target(s) being in the same pre-designated category and participants would utilise this relatedness more efficiently, providing that when perfect cue-target matching exists compared to when such blocks does not exist in the reporting sheets of the cued-recall participants. As a result, the probability of observing the expected dissociation between memory and metamemory performance in high-ITA target recall compared to low-ITA target recall was thought to be enhanced via the process of generating related targets is facilitated even further.

Lastly, it should be noted that comparing the relative merits of different methodologies to measure metacognitive processes was indeed beyond the scope of the current research. However, recently Higham (2011) compared some common methods used in the calculation of metacognitive processes (e.g., bias, control, and monitoring) such as Quantity-Accuracy Profiles (QAPs) of Koriat & Goldsmith (1996c) and type-2 signal detection theory applications (see section 3.1. of Chapter 1 for more details on these approaches). Type-2 signal detection methods indeed yield relatively more direct and accurate calculations compared to other currently available methods, such as QAP or Kruscal-Gamma (γ) correlations, which measure the correlations between correctness of the responses given and the confidence levels attached to these responses (cf. Nelson, 1984). Hence, the report type in this experiment was manipulated within participants (unlike Experiment 2), based on a report-or-pass methodology as proposed by Higham (2011; also see e.g., Higham 2002; Higham & Tam, 2005, 2006 for the application of the method). In the testing method, each participant 'report' the responses they feel that they remember or still give a response when they feel stuck in remembering by indicating that they 'pass' the response (see section 4.3.5. procedure for further details).

4.2. *Method*

4.2.1. *Participants*

One hundred and forty-three graduate and postgraduate students (age: M = 21.04, SD = 1.71) of the University of Southampton, England whose first language was English participated in the experiment. The participants were compensated for their time with course credits or £5 cash payment. Forty-two of the participants (29.4%) were male and 101 of them (70.6%) were female. The participants were randomly assigned to one of the six groups of the experiment: constrained cued-recall, liberated cued-recall, and uncued-recall groups, each of which was also manipulated with two levels of study type: pair-focused and target-focused. Table 10 displays the demographic characteristics in each group.

Table 10

Demographic Characteristics of the Participants in Experiment 3

			Age		Gender	
Study type	Group	*n*	*M*	*SD*	Male	Female
Pair-focused	Constrained CR	24	19.79	1.59	2	22
	Liberated CR	24	22.25	4.02	9	15
	Uncued recall[a]	23	23.26	3.29	13	10
	Total	71	21.75	3.42	24	47
Target-focused	Constrained CR	24	20.21	2.85	6	18
	Liberated CR	24	20.96	2.67	7	17
	Uncued recall	24	21.04	4.91	5	19
	Total	72	20.35	3.60	18	54

Note. CR = cued recall.

[a] It was suspected that a participant in the target-focused uncued-recall group misunderstood the instructions at the time of testing. This participant was detected to report the majority of left-hand side words (cues), although the targets were instructed to be the right-hand side words in the pairs. Therefore, this participant was removed from the analyses.

4.2.2. *Materials*

4.2.3. *Study lists*

The study lists used in Experiment 2 were also utilised in this experiment (see section 2.5. for the details on the study lists' construction). However, the instructions at study and testing as well as the arrangement of cue words given at the time of testing (only valid for the cued-recall groups) were different than Experiment 2 (see procedure section 4.3.5. for the details).

4.2.4. *Experimental design*

The study used a 2(study type: pair-focused vs. target-focused) X 2(recall type: cued vs. uncued) X 2(scoring instruction: constrained vs. liberated) X 2(inter-target association level: high vs. low) X 2(report type: free vs. forced) mixed design, with study type, recall type, and scoring instruction manipulated between participants and with inter-target association level and report type manipulated within participants. The dependent variables were memory quantity (MQ), monitoring performance (*d'*), which was measured on the basis of type-2 signal detection theory, and confidence level given for each response.

4.2.5. *Procedure*

Participants were tested individually in a quiet and dimly lit research station located in the School of Psychology of the University of Southampton, England. The study had two study-test cycles. Study phases were computerized and the test phases were paper-pencil. Each participant read and signed a written informed consent form before the experiment started.

In each study phase, participants were presented with a total of 20 word pairs on a Macintosh computer via Runtime Revolution software computer program. Each word pair appeared consecutively on the computer screen for 3 seconds with 1-second ISI. The pairs were located in the middle of the screen, separated with a hyphen and presented in capital letters (i.e., MUSCLE – BOOK). The participants started the presentation of the lists when they felt ready via clicking on a button located on the computer screen and to start the presentation of the pairs. The presentation of the pairs was totally randomized. Additionally, the presentation order of the study lists was counterbalanced. The participants in the pair-focused study condition were not informed which word in the pair would be the target (to-be-remembered) word. On the other hand, the participants in the target-focused study condition were provided with the information about the places of words in advance of studying lists (e.g., left hand side words in the pairs are the cue words, right hand side words are the target words). However, none of the participants were informed whether they would be provided with the cues or not at testing. Instead, they were instructed to attend to all of the words in the list as they would be asked to try and remember as many words as possible after the presentation of the list was completed – or they would only be responsible for the words on the right hand side in the pairs, targets. The target-focused participants, therefore, were reminded not to ignore the cue words since they would be helpful in remembering the target words.

Following each study phase, the participants were given two-paged reporting sheets that were specific to the conditions of the participants. The first pages of the reporting sheets involved the instructions explaining the way of reporting the responses and the second pages of the reporting sheets were designed to let the participants write down their responses. On these reporting pages, there was a column of 'CUES' involving the cue words and a 'TARGETS' column, which had empty spaces if the participant was in a cued-recall group. More importantly, the cue words were arranged in such a way that if a perfect match between the cues and the targets is made, the targets belonging to particular categories come after another and they constitute four category blocks together. On the other hand, if the participant was in an uncued-recall group, there was no 'CUES' column. Instead, there were only 20 empty spaces one under another (see Appendix C).

Regardless of the condition to which the participant was allocated, all of the participants were always asked to report 'only the target words', which were the words on the right hand side in the pairs. The constrained-cued-recall participants were instructed that their responses would be counted correct *only when* they match the targets with the cue words paired with at study. On the other hand, the liberated-cued-recall participants were instructed that they could write down the target words *without* trying to match the target words with the same cue words paired with at study. The participants had no time limit for reporting and could start at whichever target word they wished to write down without considering the presentation order.

In order to measure free-report and forced-report performance, the 'report-or-pass' method was used. The participants were asked to fill in all of spaces in the 'TARGETS' column even if they needed to make pure guesses. However, they were asked to put a tick in the 'REPORT' checkbox if they felt that the word was a target word. Alternatively, they were asked to put a tick in the 'PASS' checkbox when they were not able to accurately remember the targets, although they still needed to give a response (even if they had to make pure guesses). This reporting method yielded the numbers of correct and incorrect responses, which were either reported or passed (withheld). The responses given with the 'REPORT' option were taken for the calculation of free-report performance. On the other hand, the responses given with 'REPORT' option together with the responses given with 'PASS' option constituted forced-report performance.

The participants were also asked to rate their confidence level for each word they reported. This rating indicated how confident the participant was on the correctness of the response he/she gave. Just like Experiment 1 and 2, the participants rated their confidence levels on a Likert-type scale ranged between 1 and 7 (1 = "Not at all confident correct", 4 = "Fairly confident correct", 7 = "Completely confident correct"). The study was terminated after the second test phase of the second list was completed. Before leaving the cognitive laboratory, the participants were provided with a written debriefing statement and the researcher responded to their possible enquiries. Being self-paced at the time of testing, the study lasted between 30 and 45 minutes.

4.3. Results

4.3.1. *Memory quantity in cued recall*

The scoring instruction was manipulated only between the cued recall participants, the constrained and the liberated cued-recall groups. Therefore, the nature of the experimental design did not let us conduct a full ANOVA on memory quantity comparing the cued and the uncued recall groups with equal sample sizes. Additionally, as documented in Experiment 2, participants tend to withhold responses under a free-report option (because of stringent report criterion). That is, 'the actual retrieval performance' of the participants, which approximates what is held in the storage, is observed much clearer under forced-report rather than free-report performance. Therefore, only forced-report memory quantity of the participants were analysed so as to investigate the actual retrieval performance. Also, liberal scoring was adopted for all analyses in Experiment 3.

In order to investigate retrieval performance in cued-recall groups, a 2(study type: pair-focused vs. target-focused) X 2(instruction type: constrained vs. liberated) X 2(inter-target association level: high vs. low) mixed ANOVA was conducted on forced-report memory quantity; see Figure 7. The results showed that study type and instruction type main effects were not significant, both Fs<1. Also, there was not a significant interaction effect between study type and instruction type. However, there was an ITA main effect, $F(1, 92) = 8.906$, $p = .004$, $\eta^2 = .088$. The cued-recall participants had greater memory performance in high-ITA target recall ($M = 61.0$, $SE = 2.6$) than in low-ITA target recall ($M = 52.7$, $SE = 2.9$). That is, participants took advantage of the inter-target association at retrieval when there existed such association between target items. The results did not show any interaction effects between ITA and study type, between ITA and instruction type, and between ITA, study and instruction types, all Fs<1.[12]

[12] The same ANOVA statistics was applied to strictly scored forced-report memory quantity. The results revealed exactly the same patterns as when liberal scoring was adopted, except ITA level did not show any significant ITA main effect, $F(1, 92) = 2.066$, $p > .05$.

4.3.2. *Memory quantity in uncued recall*

A 2(study type: pair-focused vs. target-focused) X 2(inter-target association level: high vs. low) mixed ANOVA was conducted to investigate forced-report MQ in uncued-recall participants. The results revealed a study type main effect, $F(1, 45) = 19.088$, $p< .001$, $\eta^2 = .298$: the target-focused uncued-recall group had significantly greater retrieval performance ($M = 62.5$, $SE = .03$) than the pair-focused uncued-recall group ($M = 45.9$, $SE = .03$). This result was important to reveal the fact that when it was encouraged, target-focused way of encoding resulted in better memory most probably because the participants had a deeper relational encoding of the targets with each other (see e.g., Keister, 1972). Furthermore, ITA level main effect was significant, $F(1, 45) = 20.041$, $p< .001$, $\eta^2 = .308$. Uncued-recall participants had significantly higher retrieval performance in recalling categorically related targets compared to categorically unrelated targets ($M = 61.4$, $SE = .03$; $M = 47.0$, $SE = .02$, respectively). Lastly, the results did not reveal any interaction effect between study type and ITA level on forced-report memory quantity of the uncued-recall participants, $F<1$.

4.3.3. *Monitoring performance*

In order to investigate monitoring performance, a 2(study type: pair-focused vs. target-focused) X 2(instruction type: constrained vs. liberated) X 2(inter-target association: high vs. low) mixed ANOVA was conducted on Area Under the Curve (AUC) scores amongst cued-recall participants.

The calculation of AUC scores instead of *d'* was preferred because of the following reason. In calculating *d'*, some hit and false alarm rates yield values of ‘1’ or ‘0’. MacMillan and Creelman (2005) suggested that undefined *d'* values could be eliminated by replacing false alarm and/or hit rate values of ‘1’ into ‘1 - (1/2n)’, and the values of ‘0’ could be replaced by ‘1/2n’, in which ‘n’ equals to the number of observations that the rate is based on. The *d'* scores in all of the experiments were calculated, however, it appeared that there existed various cases that needed to be corrected on the basis of MacMillan and Creelman’s suggestion. In order to gather healthier scores of monitoring performance which are not much affected by this correction factor, Area Under the Curve (AUC) scores were calculated and reported in all of the experiments instead of *d'* scores (AUC is a non-parametric measurement that does not depend on the assumption of equal-variance distributions as *d'* calculations).

The results showed that there was not any significant study type and instruction type main effects, both Fs<1. The interaction effect between study type and instruction type was, however, significant, $F(1, 87) = 4.255$, $p = .042$, $\eta^2 = .047$. Pair-wise mean comparisons showed that the interaction appeared because liberated cued-recall participants monitored their responses better than constrained cued-recall participants when they studied the pairs ‘target-focused’ way at a marginally significant *p* value (*M* = .88, *SD* = .11; *M* = .80, *SD* = .20), $t(43) = 1.836$, $p = .70$, whereas they did not differ in terms of monitoring performance when they studied the words ‘pair-focused’ way (*M* = .81, *SD* = .17; *M* = .87, *SD* = .15), $t(44) = 1.153$, $p > .05$. Inter-target association level (ITA) indicated a significant main effect, $F(1, 87) = 7.113$, $p = .01$, $\eta^2 = .076$: Cued-recall participants monitored their responses in low-ITA target recall better (*M* = .89, *SD* = .02) than in high-ITA target recall (*M* = .79, *SD* = .03). That is, when targets were interrelated, this resulted in a poorer monitoring performance at retrieval –although interrelatedness amongst targets facilitated retrieval performance. Lastly, the interaction effects between ITA and study type, ITA and instruction type, both *F*s<1, and between ITA, study type, and instruction type were not significant, $F(1, 87)= 2.647$, $p>.05$.[13]

Second, monitoring performance (Area Under the Curve) amongst the uncued-recall participants was investigated. For the analysis, a 2(study type: pair-focused vs. target-focused) X 2(inter-target association level: high vs. low) mixed ANOVA was conducted. The results indicated a marginally significant main effect of study type, $F(1, 45) = 3.223$, $p = .08$, $\eta^2 = .067$. The target-focused uncued-recall participants had significantly better monitoring performance (*M* = .84, *SE* = .03) than pair-focused uncued-recall participants (*M* = .82, *SE* = .02). The results did not reveal a significant ITA main effect, and the interaction effect between ITA level and study type was not significant, both *F*s<1

In order to see the possible dissociative patterns between forced-report memory and monitoring performance (AUC) as a function of ITA level in each group as well as the specific predictions shown in Table 9, see Figure 7 which displays the means of both memory and monitoring performance in each of the experimental group. Receiver Operating Characteristics (ROCs) in each group as a function of inter-target association

[13] The same ANOVA statistics was conducted on strictly scored monitoring performance (AUC) and the results revealed exactly the same patterns as when liberal scoring was adopted, except that the following. The interaction effect between study type and scoring instruction on monitoring performance was not significant, $F(1, 92) = 2.049$, $p> .05$.

level are shown in Figure 8[14]. Complimentary to ROC curves, Table 11 displays mean confidence levels and number of targets, cues, and extra-list words reported in each group.

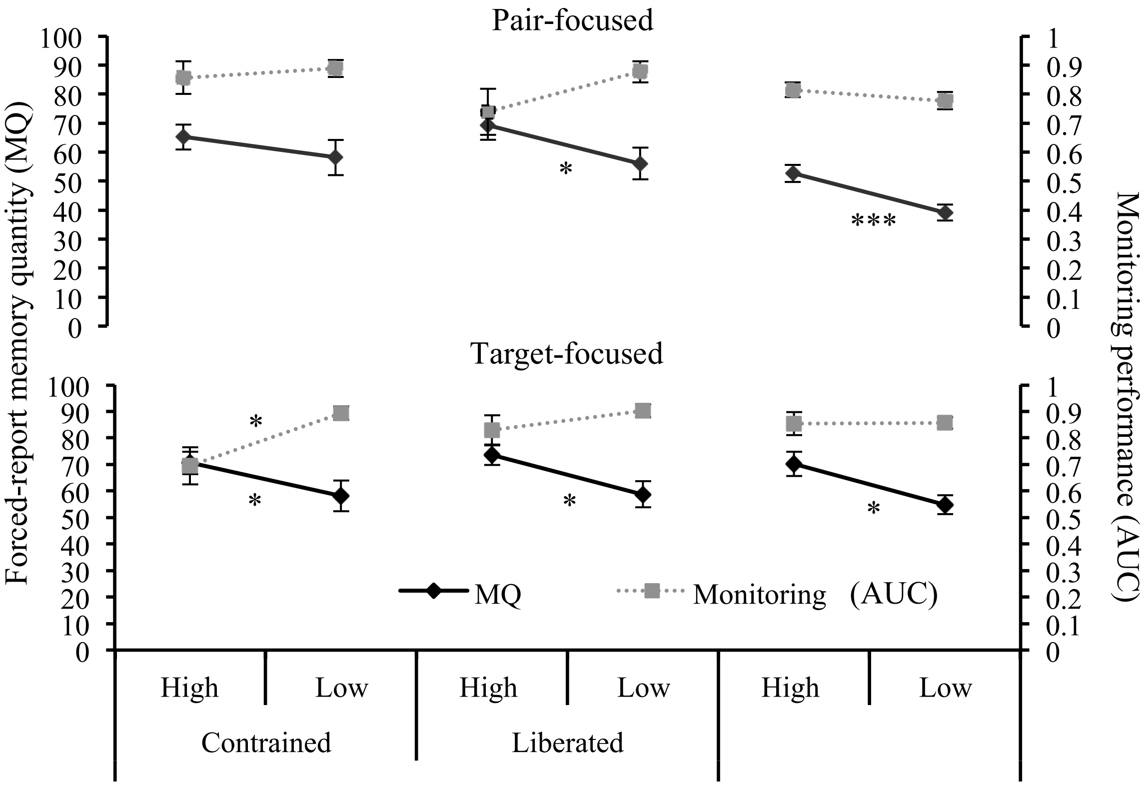


Figure 7. Forced-report memory quantity (MQ) and monitoring performance (AUC) in cued-recall and uncued-recall groups as function of types of study (pair-focused vs. target-focused), scoring instruction (constrained vs. liberated) and inter-target association level (high vs. low). Standard errors are shown with error bars attached to each score.

* = $p < .05$; *** = $p < .001$

[14] To construct ROC curves, hit rates (number of correct reported / (number of correct reported and withheld; HR)) and false alarm rates (number of incorrect reported/ (number of incorrect reported and withheld; FAR)) were calculated for each participant. These rates were shown throughout cumulative confidence levels on ROC curves; see Figure 10. The cumulative confidence levels were: 2+, 3+, 4+, 5+, 6+, and 7. The '2+' refers to the responses given any of the confidence level, ranging between 2 and 7, and '3+' refers to the responses given any of the confidence levels which ranged between 3 and 7, and so on. The highest points in ROC curves are the conjunction points of mean HR and FAR at 2+ confidence level. The lowest points are the conjunction points of mean HR and FAR which considered the responses that were given the confidence level 7. Hence, the points in between are the conjunctions of mean hit and false alarm rates at 3+, 4+, 5+, and 6+ cumulative confidence levels, from highest to the lowest points.

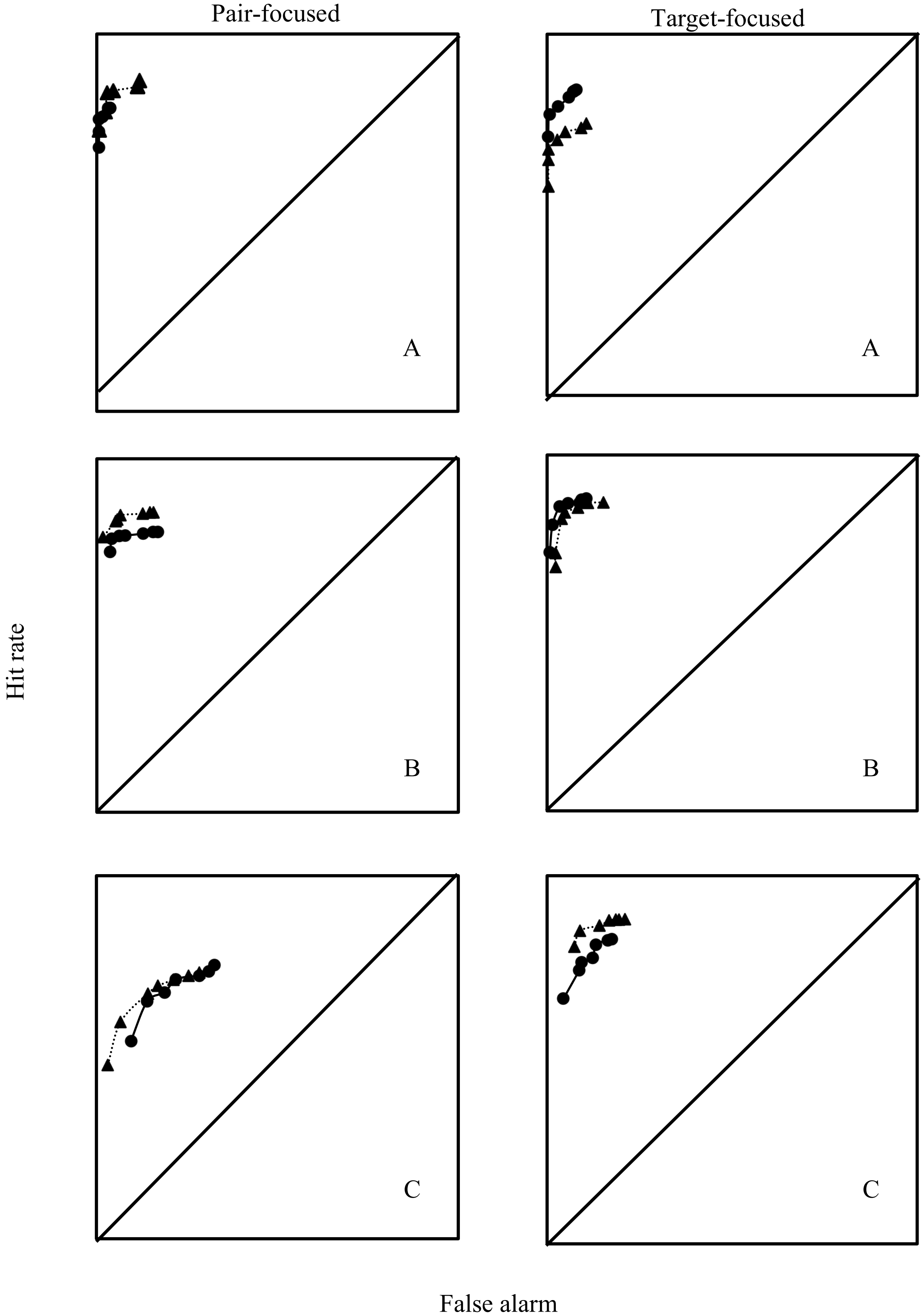


Figure 8. ROCs in constrained cued-recall (A), liberated-cued recall (B), and uncued-recall groups (C) as function of study type (pair-focused vs. target-focused) and inter-target association level (high vs. low; shown with straight and dashed lines, respectively).

Table 11

Mean Confidence Levels (CF) Targets, Cues, and Extra-list Words and the Number (Nbr) of Cue and Extra-List Intrusion Errors Reported or Passed As a Function of Inter-Target Association Level (High vs. Low), Study Type (Pair-focused vs. Target-focused) and Recall Type (Constrained Cued Recall vs. Liberated Cued Recall vs. Uncued Recall)

			Reported					
	Recall		Target		Cue		Other	
Study type	type		High	Low	High	Low	High	Low
Pair-fcs	CCr	CF	5.3 (1.4)	6.2 (0.9)	4.7 (1.2)	3.8 (0.4)	3.0 (1.3)	4.4 (2.1)
		Nbr			0.2 (0.5)	0.1 (0.4)	0.4 (0.8)	0.3 (0.7)
	LCr	CF	5.6 (1.0)	6.6 (0.7)	5.2 (1.2)	3.7 (0.8)	3.5 (1.9)	3.6 (2.1)
		Nbr			0.2 (0.6)	0.5 (0.8)	1.0 (1.9)	1.2 (1.7)
	UncR	CF	5.0 (1.3)	5.7 (1.0)	4.3 (1.3)	4.7 (1.4)	3.4 (2.2)	3.3 (1.3)
		Nbr			1.6 (1.6)	2.3 (1.8)	1.2 (2.3)	1.8 (2.3)
Target-fcs	CCr	CF	4.9 (1.6)	6.1 (1.1)	2.0 (-)	-	3.0 (1.9)	4.2 (1.6)
		Nbr			0.0 (0.2)	0.0 (0.0)	0.5 (0.8)	0.7 (1.3)
	LCr	CF	4.9 (1.7)	5.8 (1.3)	-	4.3 (3.1)	3.5 (1.9)	3.1 (1.2)
		Nbr			0.0 (0.0)	0.1 (0.3)	0.8 (1.6)	1.0 (1.9)
	UncR	CF	5.4 (1.5)	6.0 (1.2)	4.0 (1.8)	4.7 (2.0)	4.8 (2.0)	3.1 (1.4)
		Nbr			0.8 (1.2)	0.7 (0.8)	0.6 (1.1)	0.6 (0.9)
			Passed					
			Target		Cue		Other	
			High	Low	High	Low	High	Low
Pair-fcs	CCr	CF	5.4 (1.2)	6.2 (0.9)	1.2 (0.5)	1.4 (0.5)	1.2 (0.3)	1.1 (0.2)
		Nbr			0.7 (0.9)	0.9 (1.2)	6.3 (3.8)	7.5 (5.3)
	LCr	CF	5.5 (1.0)	6.6 (0.6)	2.2 (1.7)	1.2 (0.4)	1.4 (0.5)	1.3 (0.5)
		Nbr			0.7 (1.2)	1.2 (1.8)	5.5 (4.5)	7.6 (4.5)
	UncR	CF	4.5 (1.1)	5.0 (1.2)	1.4 (0.6)	1.6 (0.9)	1.4 (0.7)	1.4 (0.6)
		Nbr			4.0 (1.4)	4.8 (2.2)	5.4 (3.3)	7.4 (3.8)
Target-fcs	CCr	CF		6.1 (0.9)	1.0 (0.0)	1.6 (0.5)	1.1 (0.3)	1.1 (0.2)
		Nbr			0.3 (0.6)	0.3 (0.6)	5.5 (3.9)	8.1 (5.7)
	LCr	CF	4.8 (1.8)	5.9 (1.2)	1.0 (0.0)	1.2 (0.5)	1.1 (0.2)	1.3 (0.6)
		Nbr			0.2 (0.4)	0.8 (1.1)	5.1 (3.6)	7.5 (4.7)
	UncR	CF	4.8 (1.6)	5.6 (1.2)	2.2 (1.4)	1.4 (0.7)	1.3 (0.5)	1.3 (0.5)
		Nbr			2.0 (2.1)	3.1 (2.3)	4.0 (3.1)	5.9 (3.3)

Note. Pair-fcs = paired-focused group; Target-fcs= target-focused group; CCr = constrained-cued-recall group; LCr = liberated-cued-recall group; UncR= uncued-recall group. Standard deviations are shown in parentheses. In order to see, percent recall (%) of the target words, see Figure 7.

4.4. Discussion

The present study aimed to investigate the effect of inter-target association in cued and uncued recall on memory as well as metamemory performance as a function of variations in the study and retrieval contexts. These manipulations were introduced so as to make specific predictions regarding the conditions in which high inter-target association is expected to be utilised in remembering the south-after information, thus possible dissociation between memory and metacognitive monitoring could be expected.

First, the results showed that the facilitation in the retrieval of target items due to high inter-relatedness between t-b-r items was observed clearly in uncued-recall participants. This result was consistent with some earlier findings that inter-target association (e.g., categorical relationship) yields better memory performance in free (uncued) recall (e.g., Tulving & Pearlstone, 1966; Puff, Murphy & Ferrara, 1971). As expected, regardless of the type of encoding (pair-focused or target-focused strategies), the uncued-recall participants retrieved higher percentages of categorically related targets compared to categorically unrelated targets regardless of the type of encoding (paired-focused or target-focused strategies). This result was attributed to the implication that the existence of inter-target association in the study list facilitated the generation process at retrieval. Based on the generate-recognize understanding, the existence of inter-target association between targets seemed to restrict the search set as long as categorical relatedness was better realised so that higher-order unit(s) were formed accordingly. Moreover, semantic association between to-be-remembered items eased the retrieval of other interrelated targets when at least one related target was remembered, 'even if' higher-order units could not have been formed. It was also interesting to observe that when uncued at testing, studying pairs in a pair-focused way also enhanced retrieval performance of categorically related targets similar to the target-focused way of encoding. In other words, this observation showed that the uncued-recall participants took the advantage of inter-target association at recall (when such an aid existed in the study lists) regardless of the way that they encoded the pairs. This was also important to show that when uncued, the participants seemed to seek any available source (in this case, ITA) to retrieve the target items rather than trying to remember the cues to help retrieval of the targets with them.

The reasons for higher retrieval performance of categorically related targets compared to unrelated targets observed amongst the uncued-recall group could, however, be attributed to different organizational effects, conceptual or associative, depending on the encoding strategy utilised. In a strict way of thinking, for instance, if no possibility of constructing a conceptual (categorical) organization among high-ITA targets when uncued-recall participants study pairs as pairs is adopted, then the facilitated retrieval of categorically related targets might be attributed to the mere elicitation of one target to the other related targets (associative organization).

The possibility of elicitation of one target to the other categorically (as well as semantically) related targets in free recall was supported by some earlier findings. Hudsin and Austin (1970), for instance, tested recall performance of their participants who studied a list of 30 words, of which every three words were from a particular experimentally designated category. The participants were tested in one of three recall conditions. In the *context group*, one word from each category was given to the participants and the remaining words in the same category were asked to be reported. In the *category group*, the participants were given category names and the words studied were asked to be recalled and no other cue words (e.g., an exemplar from the studied category) were given to the participants. In the *free-recall control group*, however, the participants were asked to remember as many words as possible. Both context and category groups recalled more words than the control group and the two former groups did not differ from each other (Experiment 1). In Experiment 2 of Hudsin and Austin, however, the category group exceeded the context group (as well as control group) in terms of the number of words recalled and there was no difference between the latter two. Hudsin and Austin suggested that the discrepancy that occurred between Experiments 1 and 2 meant that the words in the study list used in Experiment 2 had relatively weaker direct associations between each other so as to trigger the retrieval of other targets. Therefore, they reasoned that the context group might have had a tendency to retrieve other targets by elicitation, if category unit is not realised or used (Hudsin & Austin, 1970). In a review, Battig (1966) also pointed out an observation from his previous studies on paired-associate learning (e.g., see Battig, 1964), which suggested that participants tend to group two or more pairs together at recall even after one-trial learning. He further stressed that no matter what measures are taken to prevent or discourage the use of grouping the pairs in terms of chunks, clusters, etc., the learning of paired-associates are remarkably persistent about grouping or coding various pairs together. Based on these findings, unlike target-focused uncued-recall group, the pair-focused uncued-recall group in the current study might have formed a semantic association between targets, even if they could not construct a higher-order unit (e.g., conceptual organization).

Experiment 3, however, showed that the target-focused uncued-recall participants retrieved even higher percentage of high-ITA targets (M = 70.2, SD = 22.36) compared to pair-focused uncued-recall groups (M = 52.6, SD = 13.97), $t(45)$ = 3.222, p = .002. The finding implied that the facilitated retrieval of targets having categorical association was even further enhanced with target-focused encoding. Therefore, the target-focused uncued-recall participants seemed to be able to construct categorical organization(s) by which they could cue themselves to guide their recall, even if the pair-focused uncued-recall group could not have done so. It should be noted that the guidance in searching the sought-after information, which was expectedly triggered by the higher-order units, was not considered as something completely eliminating the possibility that retrieval of one target might elicit the activation of another related target. In other words, conceptual organization was not considered to be as an 'either-or' process, whereas associative organization might be a factor facilitating the retrieval of related targets via fundamentally –even if not only- activating other items with retrieval of related targets on the basis of semantic relatedness between them (this possibility will be examined in Experiment 6; also see section 6.5.. for further elaboration on the subject). The elicitation of a target activated by the other related items remembered mainly converges with the *spread of activation hypothesis* (Collins & Loftus, 1975) incorporated clearly in the *associative network model* of Bower (e.g., 1980). According to the model, memory consists of clusters of nodes where words, concepts and events (among others) are represented and related items are assumed to be connected to each other. As a result, when memory of an item is accessed and activated, it is believed that the energy spreads along these connections so that the activation of the related items is facilitated (Bower, 1980).

Second, the results on cued recall showed that the retrieval-time existence of the cues encoded with the t-b-r item was a quite effective factor in enhancing recall performance. However, it was found that the target-focused cued-recall participants had a higher percentage of mismatched targets (observed in the difference between strictly and liberally scored retrieval performance) compared to pair-focused cued-recall participants. This result implied that when the participants study pairs focusing on targets more, they indeed could not utilise the cue existence at test as much as pair-focused cued-recall participants could. That result was in line with the expectations of encoding specificity principle (Thomson & Tulving, 1970) in terms of revealing that cues given at retrieval are only (or more) effective when they were encoded with the to-be-remembered item at study (e.g., Craik & Tulving, 1975). However, it should be noted that the encoding specificity principle herein is interpreted in the following way. The higher the cue was incorporated with the target, the higher possibility of that cue to elicit the target encoded together. As Nairne (2002) argued, however, it is believed that it is not the sheer availability of the cues at testing which is sufficient for an effective retrieval performance, it is the relative strength of the cues to prompt the target retrieval (see section 1.5.3. in Chapter 1 for the details on the argument). In other words, the more the cue is 'overloaded' with various 'related targets', the less possibility of that cue to elicit or activate any particular target encoded with that cue item. In the current experiment, however, it was not the cues but the targets that turned out to be less distinctive to be elicited by their paired cues. Hence, the percentage of mismatched targets in target-focused cued-recall participants was greater compared to pair-focused cued-recall participants. This finding also took the support from the levels of processing approach as well (Craik & Lockhart, 1972) in the way that target-focused way of encoding resulted in a lowered power of the cues to ease retrieval of their paired targets compared to the cues encoded with pair-focused encoding strategy (indexed by the higher mismatched targets in target-focused cued-recall group compared to pair-focused cued-recall group).

More specifically, the cue availability at recall was found to be equally effective when cue-target matching was considered (strict scoring), regardless of whether the inter-target association existed or not. On the other hand, when cue-target matching was ignored at scoring (liberal scoring), cued-recall participants had higher retrieval performance for categorically related targets compared to categorically unrelated ones. In short, facilitated retrieval via target association was observable only when cue-target matching was ignored. This was important to reveal the fact that the cued-recall participants depended more on the cue existence at recall rather than target association, as well as the fact that utilising inter-target association to aid retrieval is observable when responses were liberally scored. This observation seems to converge with the idea that the generate-recognize route operates when direct access fails (Bahrick, 1969, 1970; see also Higham & Tam, 2005). The cued-recall participants seemed to report their responses on the basis of cues (observed by comparable recall percentages for related and unrelated targets, which were matched correctly), which seems that they utilised the direct access – even if not solely. However, when available, they could utilise inter-target relatedness to enhance retrieval via the generate-recognize route most probably after direct access started to fail at some point, which in turn resulted in a higher number of mismatched responses.

Third, the results indicated that when target association was engaged so as to enhance memory, monitoring performance (e.g. *d'* scores measured on the basis of type-2 signal detection theory) for categorically related target recall could not be facilitated accordingly. Dissociation between memory and metamemory performance was observed particularly in uncued recall regardless of the study type and in the cued-recall participants who studied the pairs in target-focused way. Even though the retrieval performance for related targets was enhanced compared to unrelated targets, monitoring in the unrelated targets recall was better compared to the related targets in cued recall. This pattern was, however, observed much more clearly amongst the target-focused cued-recall participant than the pair-focused cued-recall participants. The different patterns observed between pair-focused and target-focused cued-recall participants further showed that the inter-target association was taken as retrieval help when it was realised and formed at the time of study[15].

The interaction between study and retrieval contexts was found to be an important factor in recall performance enhanced via inter-target association, thereby affecting monitoring performance not in the parallel way. The results confirmed all of the expectations specified in Table 9; see Figure 7. The only exception was the expectation concerning the pair-focused liberated-cued-recall group when liberal scoring was adopted at counting the correct and incorrect responses. It was shown that when liberated from cue-target matching, pair-focused cued-recall participants had higher retrieval as well as higher monitoring performance in high-ITA target recall compared to low-ITA target recall. This pattern was attributed to the following. These participants had some mismatched targets by taking the advantage of ITA level. However, as indexed with higher monitoring performance in high-ITA target recall compared to low-ITA one, these mismatched responses did not seem to be, for instance, extra-list items (e.g., extra-list category exemplars) just like target-focused liberated-cued recall group. In other words, adopting a pair-focused encoding seemed to hinder having intrusion errors as much as possible, even though it yielded mismatched targets when being liberated at retrieval. Further, it was found that when the constrained cued-recall participants were asked to study the pairs as pairs, they had comparable retrieval and monitoring performance for high and low-ITA targets. Whether targets were categorically related or not did not seem to be an essential factor in the constrained-

[15] Amongst some others, the question as to whether the effect of inter-target association on enhancing retrieval is effective only when it is formed at encoding or it might be formed at retrieval will be addressed in Experiment 6.

cued-recall who used pair-focused encoding to affect memory, in turn metamemory performance. However, the constrained-cued-recall group who studied the pairs by focusing on the target words more (by which the targets were expectedly less connected with the cue words), the advantage of the categorical relationship amongst targets on retrieval came to the stage. This result seems to explain why organizational effects are observed clearly in the uncued (free) recall testings (e.g., Tulving, 1962; Puff, Murphy & Ferrara, 1971) rather than conventional pair-associate learning experiments. Because, both of the findings implied that when cue-targets are well encoded together (e.g., by pair-focused encoding), no other variables such as the interrelatedness between t-b-r items, or being constrained to make cue-target matching or being liberated to do so at retrieval do not make a difference over cue help. More importantly, however, the results showed that retrieval performance did not facilitate the monitoring performance in the parallel fashion in target-focused cued-recall participants, regardless of the scoring instruction (partial dissociation).

The patterns mentioned so far imply that the possible dissociative effect of the ITA level emerges depending on the effectiveness of the encoding processes, which let forming associations between to-be-remembered material at the time of study. In other words, in conjunction with the observation of Battig (1966; see also, 1964) who pointed out the tendency of participants to group the pairs together even though they were discouraged to do so, cue dependency in retrieval reducing as a result of studying pairs via target-focused way seemed to allow the cued-recall participants to take advantage of categorical relatedness when such associations existed (even further in uncued recall).

In short, the comparisons made between retrieval and monitoring performance as a function of ITA level showed that the way of the encoding strategy utilised by the participants work as a factor to repress or reveal the utilization of categorical relatedness between targets for retrieval particularly in cued-recall, after firstly depending more on the cues when they are available. Additionally, liberating the cued-recall participants in terms of cue-target matching at the time of testing work as a facilitative factor in retrieval which emerge with a target-focused encoding strategy when studying targets (observed with liberal scoring). The following experiment, which aimed to show a clear dissociation between memory and monitoring performance as a function of the ITA level in the study lists, therefore, utilised only target-focused way of encoding. In other words, as in conventional paired-associate learning experiments, the participants would be informed in advance that which words were to be cues and which words were to be to-be-remembered items at testing.

CHAPTER 5
Experiment 4[16]

5.1. Introduction

It has been suggested that retrieval of to-be-remembered (t-b-r) items in cued recall can be achieved via two processes: an efficient 'direct' retrieval process, in which the details of the to-be-remembered item (e.g., ecphory; cf. Tulving, 1982) is recollected and a process that operates when direct access fails and involves covert 'generation' of possible candidates preceded by an attempt to 'recognize' the t-b-r item amongst the candidates. The latter process underlies the generate-recognize models (e.g., Kintsch, 1968; Bahrick, 1969, 1970; Anderson & Bower, 1972). However, recognizing the target information amongst the alternatives given, such as in conventional recognition tests (e.g., Benjamin, 2003; Benjamin & Bawa, 2004), or amongst highly related candidates was suggested to be difficult (Higham & Tam, 2005; see also Higham, 2002). In the context of self-generated alternatives, for instance, Higham and Tam (2005) investigated memory and metacognitive performance of cued recall. Unlike weak semantic associate cue-target pairs (i.e., 'shampoo-blood'), the strong semantic association in the pairs (i.e., 'vein-blood') could be facilitative in the process of generating and retrieving the target item ('blood') when the participants are asked to recall the target using a cue (e.g., 'vein-?). However, the possibility of recognizing the correct response (target) amongst generated candidates could be reduced since the generated responses are highly probable to be a related item (Higham & Tam, 2005).

In relation to the recognition difficulty of generated candidates having common features in some dimension (i.e., categorical, semantic, and/or structural relationships), the studies on false memory (e.g., Kato, 1985; Roediger & McDermott, 1995; Roediger, Watson, McDermott, & Gallo, 2001; Kelly & Sahakyan, 2003) and the distinctiveness effects on memory (e.g., Schacter, Israel & Racine, 1999; Gallo, Weis, & Schacter, 2004) seem noteworthy. For instance, Kato (1985) developed an interesting cued-recall paradigm in which many false recalls have been 'produced' by participants who study a set of paired words, half related and half unrelated. The target word, which is always

[16] This study was presented in Medicine, Health, and Life Sciences Postgraduate Conference on 10th June 2010 in Southampton, England, United Kingdom; and, together with Experiment 6, it was presented in 51st Annual Meeting of the Psychonomic Society on 19th November 2010 in St Louis, Missouri, United States of America.

the second word in the pair (e.g., clock – dollar), is cued with the context word as well as three letters of the to-be-remembered item at the time of testing (clock – do ___ r). However, some of the unrelated word pairs are deceptive in the way that an 'unstudied' associate of the context word (i.e., nurse) fits in target recall since the three letters given as part of the cue are exactly the same as of the studied target (e.g., nurse – do ___ r; in this case, the competing target is 'doctor'). The results of Kato (1985) showed that participants produced the strongly associated competitor (e.g., 'doctor') almost as often as they recalled the studied item ('dollar'). Kato (1985) suggested that the cued competitor was accessed so fluently and easily that participants considered that it had actually been studied. Kato considered this fluency as a form of *fluency heuristic,* suggested by Jacoby and Hollingshead (1990; see also Jacoby & Dallas, 1981; Jacoby, Kelley, & Dywan, 1989a; Lindsay & Kelley, 1996). More importantly, however, the false recalls created in Kato's paradigm were found to affect monitoring negatively (Kato, 1985). In other words, participants were poorer in their ability to discriminate between their correct responses and incorrect ones. Kelly and Sahakyan (2003) also used Kato's cued-recall paradigm. They found that the deceptive items resulted in lower *monitoring effectiven*ess compared to the related and control (unrelated) cues (Kelly & Sahakyan, 2003). The monitoring effectiveness, in this vein, was borrowed from Koriat and Goldsmith's model on strategic regulation of memory accuracy (1996c). It mainly referred to the degree to which assessed probabilities of correctness (P_a) successfully differentiate between correct and incorrect candidate answers (see section 1.3.1. and Figure 2 for more details on the framework of Koriat & Goldsmith, 1996c).

False memory has also been studied in recognition. Israel and Schacter (1997), for instance, tested recognition performance of a group of participants. One group of participants were presented with a list of semantically-related words in auditory format and each word was accompanied with a picture. Another group of participants, however, studied the same words that were presented in auditory format, however, each word was accompanied by a visual representation of the word instead of being accompanied by pictures unrelated to the studied words. Israel and Schacter showed that the study items encoded with pictures had lower false recognition in both related and unrelated new items (lures) than the words encoded with their own visual presentations. They proposed that participants use a process called *distinctiveness heuristic* (Schacter, Israel, & Racine, 1999) that refers to a retrieval monitoring process in which recollecting the distinctive information is used so as to eliminate false recognition.

The effect of distinctiveness of the studied material on retrieval performance has also taken the interest of a considerable number of scholars since Calkins' influential studies on the *isolation paradigm* (1894, 1986). In this paradigm, participants are presented with material to be learned and a small proportion of it differs on some dimension from the majority of the material. The result of this isolation is enhancement of memory for the different material (e.g., see Hunt, 2006 for a review). As a clear example of the distinctiveness effect, Benjamin and Bawa (2004; see also Benjamin, 2003) demonstrated the following. Benjamin and Bawa basically investigated criterion placement in recognition memory of participants who were tested in two successive recognition tests (e.g., 'old/new' judgements) as a function of variations in the *distracter plausibility*. The distracter plausibility referred to the degree to which new items (foils) have some commonalities, such as categorical relatedness, with the old items. The results of Benjamin and Bawa showed that the participants shifted to a more conservative criterion on a second test when it involved more plausible old items (overlapping in terms of categorical association) than the prior test. On the other hand, they found no evidence of a criterion change when the second test involved less plausible distracters compared to the previous one. Benjamin and Bawa inferred that the shift was triggered, at least partially, by the assessment of actual performance. They reasoned that the setting of criterion in recognition was supported by the assessment of discriminability between old and new items instead of the pre-test probabilities of studied and unstudied items or memory strength of the studied items (Benjamin & Bawa, 2004). The conclusion of Benjamin and Bawa, along with the argument of Higham and Tam (2005) underlined the main prediction of the current experiment: when the association (e.g., categorical relationship) between target items is high in paired-associate learning, in which the cue-target pairs are weak associates, recognition of a target would be difficult amongst generated candidates which have expectedly comparable memory strengths to each other, so that they would have a high chance of being the target. As a result, as long as memory performance for the target recall could be enhanced via high-ITA, monitoring of the correct responses (e.g., *d'* scores measured in terms of type-2 signal detection theory) would conversely be lowered.

5.2. The aims and expectations of Experiment 4

The experiments reported so far empirically support the predictions made on the basis of generate-recognize approach (e.g., Bahrick, 1970; Kintsch, 1970; Anderson & Bower, 1972). Experiments 1 and 2 confirmed that participants generate more correct responses than they report, since they tend to withhold some of the correct responses when a stringent report criterion (free-report option) is employed. In better relation to the aim of the current research, Experiments 1, 2, and 3 overall showed that when the to-be-remembered (t-b-r) items have experimentally-determined inter-relationships (such as categorical association), participants have a tendency to make use of this knowledge to aid recall of the stored information. This interrelatedness between t-b-r items was considered to restrict the search set that guides the retrieval process as much as the tasks at study (e.g., study type/depth of encoding) and the task demands at retrieval time (e.g., constrained or liberated in cue-target matching particularly when pairs were studied with target-focused encoding) allow the participants to use that knowledge acquired by such as the list structure (see, Experiment 3). The importance of the tasks at study (e.g., study type) in paired-associative learning when utilising the inter-relatedness of the t-b-r material in retrieval was, however, found to be a more important factor in changing the retrieval patterns in cued recall compared to uncued recall. Therefore, the facilitative effect of categorization on retrieval that does not affect monitoring similarly was observed more clearly in uncued recall compared to cued recall. As documented in the literature on paired-associate learning and was observed in Experiment 3, inter-target associations were found more useful in uncued recall (e.g., Bousfield, 1953; Cohen, 1966; Jenkins, Milk & Russel, 1958; Tulving, 1962), since participants seem to depend on other available sources rather than trying to retrieve the cues to remember their paired target items. Moreover, the dissociation between memory and metacognitive monitoring that is observed in uncued recall is enhanced even further with the target-focused encoding strategy that is found to let participants organize the material in a better way compared to the pair-focused encoding (Experiment 3). In short, the common denominator of the findings in the previous experiments is that providing a restricted memory search guided by organizing the study material enhances memory performance but this is accompanied by a cost on monitoring (e.g., see Higham & Tam, 2005).

The above-mentioned dissociation was, however, shown with a partial dissociation in the previous experiments (e.g., Experiments 2 & 3). In this context, the *partial dissociation* referred to the following observations. Facilitated memory performance of related targets is accompanied by comparable monitoring performance as opposed to unrelated targets. Alternatively, it referred to comparable memory performance of related targets to unrelated targets that goes along with lower monitoring performance in the recall of the related targets compared to unrelated ones. Therefore, the primary objective of Experiment 4 was to reveal a *full dissociation* between memory and monitoring performance as a function of inter-target association level. By full dissociation, I mean that there should be significantly higher retrieval performance and significantly lower monitoring performance of categorically related targets compared to categorically unrelated ones.

For the purpose of showing a full dissociation between memory and monitoring performance in the inter-related target recall, the number of categories to which target items belong in the study lists was manipulated in the current experiment. Specifically, three lists involving 24 word pairs in each had the target items as exemplars of either two, six, or twenty-four categories. Considering that the number of studied pairs between lists is constant (24), the manipulation of the number of categories can also be considered a manipulation of category size. In other words, as the number of categories increases (two, six, twenty-four), the category size in each list decreases (twelve, four, one)[17].

Investigating retrieval performance of categorically and/or semantically related material with regards to the variation in study and test contexts is indeed not something new. For instance, Tulving and Pearlstone (1966) tested the recall performance for categorized lists with the primary aim of studying the accessibility of available information with appropriate cues. They tested participants' cued- and free-recall performance after the participants studied single-item lists. Further, Tulving and Pearlstone (1966) manipulated the category size (one-two-four) as well as the list length (12-24-48). The results of their study showed that cued recall, in which category names were given to prompt recall, resulted in greater recall performance than the uncued-recall performance of the same material. More importantly, Tulving and Pearlstone (1966) showed that the category names given increased the number of items reported substantially and directly with the length of the lists. However, the words recalled within each category were found to be independent of the recall conditions (cued vs.

[17] For a consistent interpretation, the variable on the structure of study lists will be considered as the manipulation on the 'number of categories' (two vs. six vs. twenty-four) across the chapter rather than a manipulation on category size (twelve vs. four vs. one).

uncued) when list length increased from 24 to 48. As a result, the researchers reasoned that whereas the probability of retrieving higher-order units (e.g., categories the items belong to) increased with the appropriate retrieval cues (category names) and the list length, accessing the items within higher order units is largely independent of these variables. However, Tulving and Pearlstone (1966) did not consider these findings as eliminating the possibility of retrieving or constituting higher-order units without any covert cues given, such as category names. In other words, even if participants were not provided with any covert cues at the time of retrieval, they might still have a tendency to construct their own higher-order units. This possibility relied on the proposal of *subjective organization* of learning material (Tulving, 1962), in which participants are found to report random (normatively unrelated) words in an order that is idiosyncratic to each rememberer.

Although providing category names as cues would expectedly boost the formation of organization in the study material, the participants in Experiment 4 were allowed to form their own higher-order units, and so no category names were provided at study or at test. This was done for two reasons. First, the probability of participants realizing the relatedness between targets in two-category lists would expectedly be higher than six-category lists, and almost definitely higher compared to twenty-four-category lists. Second, not providing category names (either at the time of study or test, or both) would give us a chance to observe participants' own retrieval strategies. Besides, participants were informed about the places of the cue and the target words in advance of testing (similar to the target-focused group in Experiment 3), which would let participants realise the inter-target associations so that they would be better able to construct higher-order units, particularly in two-category lists - and possibly in six-category lists.

As support for the expectation that participants are able to construct their own retrieval cues (e.g., which categories existed in the study lists), the results obtained in the studies of Slamecka (1968, 1972) are notable. Slamecka (1968, Experiment 1) tested recall performance of participants for three study lists, which contained 30 t-b-r items. Those words were either rare or common words or the free associates of a predetermined concept, such as 'butterfly'. Participants were tested either in a *context group*, in which half of the words studied (targets) were given to cue the remaining targets, or in a *control group*, where participants were asked to report all the words studied without any cues. Counter intuitively, the context group was significantly inferior to the control group on the critical recall measure (e.g., total presented minus context words presented). Slamecka (1972) also varied the number of categories and showed that the cueing effect was only effective as the number of categories studied increased. He reasoned that the studies which found a positive effect of cuing (such as providing category names to participants) on free recall performance (e.g., Tulving & Pearlstone; Hudson & Austin, 1970; Luek, McLauglin & Cicala, 1971) is understandable since the higher the number of categories used, the more difficult it is to form a higher-order unit by the participants when the category names are not provided.

In addition, the *temporal context model* of Howard and Kahana (1999, 2002) does not seem at odds with the expectation that participants could form higher-order units to aid retrieval as cueing themselves. According to the model, a candidate memory is thought to evoke retrieval of its temporal context in the study list. Further, Howard and Kahana (2002) argued that retrieved context information could easily serve as a retrieval cue for other list items. Taking an inference from the model, the probability of evoking the retrieval context, in which various degrees of categorical inter-relatedness between the t-b-r items exist (e.g., number of categories in the study lists), was expected to be higher in two-category target recall compared to the target recalls from six-category, and twenty-four category lists. Additionally, the observations of Battig (1966; also see Battig, 1964) that lay out the fact that participants have a tendency to form groupings of the pairs no matter what precautions are taken or whether they were discouraged to do so, also support the expectation that participants could organize the inter-related targets as long as such inter-relatedness exists in the study list.

Therefore, it was hypothesized that the forced-report memory quantity (actual retrieval performance) in two-category target recall would be higher compared to six-category as well as twenty-four-category target recalls. However, the monitoring performance in two-category target recall was expected to be lower compared to six-category and twenty-four-category target recall (full dissociation). The rationale behind these hypotheses lay on the suggestions of Higham and Tam (2005). It was predicted that high inter-relatedness between to-be-remembered items would facilitate the generation process of candidates; However, recognition of the correct response (target item) from amongst the candidates would be difficult due to the high possibility that these candidates would be highly-related to each other.

5.3. *Method*

5.3.1. *Participants*

Sixty undergraduate and postgraduate students in the University of Southampton, England participated in the study. Each participant was compensated for his/her time with course credits or £5 payment. The participants were randomly assigned to one of two groups: cued recall and uncued recall. The demographic characteristics of the groups are shown in Table 12.

Table 12

Demographic Characteristics of the Groups in Experiment 4

		Age		Gender	
Recall Type	*n /N*	*M*	*SD*	Male	Female
Cued recall	30	19.60	3.82	3 (10.0%)	27 (90.0%)
Uncued recall	30	19.23	1.48	2 (6.7%)	28 (93.3%)
Total	60	19.42	2.88	5 (8.3%)	55 (91.7%)

5.3.2. *Experimental design*

The experiment used a 2(recall type: cued vs. uncued) X 2(report type: free vs. forced) X 3(number of categories in the study lists: 2-6-24) mixed design with recall type as the between-participants variable, and report type and number of categories in the study lists as the within-participants variables. The dependent variables were memory quantity, confidence levels and monitoring performance, *d’,* which was measured on the basis of type-2 signal detection models as in Experiment 3.

5.3.3. *Materials*

5.3.4. *Study lists*

Three study lists involving 24 word pairs were created. All lists had weakly-associated cue-target pairs (e.g., ‘VICTIM – TEACHER’, ‘IMPACT – MANAGER’, etc.). However, the lists involved the target words (the words on the right hand side in the pairs) being either the exemplars of two, six, or twenty-four categories. First, the target words were determined. For this purpose, twenty-four categories were selected from 72 available categories in the category norms study of Van Overschelde et al. (2004) on the basis of two criteria: The selected categories were mutually exclusive and each selected category had at least 13 exemplars available in the norm study of Van

Overschelde et al. (2004); see Appendix D for the selected categories. Therefore, 13 exemplars for each 24 categories (a total of 156 words) constituted the 'targets pool'. Second, a total of 72 words were determined so as to be paired with the target words which involved in the three study lists (each of the lists had 24 pairs). These words constituted the 'cues pool'. The cue words were decided on the basis of a criterion that they had no semantic (as well as categorical) association between each other and with any other target words, selected from the University of South Florida Free Association Norms (Nelson, McEvoy, & Schreiber, 1998).

The lists were set up individually for each participant just before the experiment started. The decisions of which categories would be in the study lists, which words would be targets, and which words would be paired with the targets as cue words were completely random. For this purpose, Runtime Revolution software computer program was used. The program randomly selected two of the categories amongst 24 available categories (targets pool). It also randomly selected 12 exemplars from the two categories. These words are the target words for the two-category list for the participant who was about to take the study. The program again randomly selected six categories from the remaining 22 categories, as well as four exemplars from each of those six categories. These words constructed the target words in the six-category list of this participant. Lastly, the computer program took the 13th exemplar that had not already been selected for the two-category list (because two-category list had already taken twelve exemplars from these randomly selected two categories). It also randomly selected one exemplar word from each of the remaining 22 categories, which had not already been selected for the six-category list. Hence, the last selection constructed the target words in the twenty-four-category list for this participant. Based on this list-composing procedure, none of the lists (two-category, six-category, and twenty-four-category) involved the same categories and the same target words. Beside target selection, the program randomly selected and assigned every 24 words of the cues pool to each of the study lists designated to be the cue words.

5.3.5. *Procedure*

The participants were tested individually in a quiet and dimly lit cognitive laboratory located in the School of Psychology of the University of Southampton. They were tested in three computerized study-test cycles. In each study phase, the participants were presented with one of the study lists: two-category, six-category, and twenty-four-category target lists. The presentation order of the lists was counterbalanced. The counterbalancing resulted in six versions of the presentation order and the participants were randomly assigned to one of the versions.

The participants had a practice study-test cycle with a study list of 5 word pairs (which were not from the pools of the targets or the cues). After the practice phase, the participants were warned that they were about to start the actual study in which their responses would be scored. In each study phase, the participants were presented with a study list of 24 word pairs randomly and they were instructed that they would be responsible for remembering and reporting the target words: the words on the right hand side in the pairs. The participants started the presentation of the lists by clicking on a “start the presentation” button located on the computer screen when they felt ready. Each word pair was presented on the computer screen for 3 seconds, with a 1-second ISI. The pairs were presented in capital letters and separated with a hyphen between the words (i.e., ‘EFFORT – UNCLE’). Each participant solved some moderate difficulty algebra calculations or Sudoku puzzles for 5 minutes as a distracter activity just after the presentation of each list.

In each computerized-testing phase, the cued-recall participants were asked to write down the target words under the ‘TARGETS’ column by using the cues provided under the ‘CUES’ column; see Appendix D. However, the uncued-recall participants were not given any cue words and were asked to write down the target words under the ‘TARGETS’ column. They instead had 24 empty spaces, one under another, designed to write down their responses. As in Experiment 3, the participants checked the ‘REPORT’ checkbox to indicate that they felt comfortable to provide that answer. Alternatively, they checked the ‘PASS’ checkbox when they felt stuck in remembering. However, they still had to give a response, even though they chose ‘PASS’ option. Cued-recall participants could start and continue giving responses to whichever cue they wished. However, the participants also indicated ‘the order of their reporting’ (1 to 24) under the ‘ORDER’ column. Lastly, they rated each response in terms of how confident he/she was that the response was correct on a Likert-type scale provided next to each response that ranged between 1 and 7 (1 = “Not at all confident correct”; 4 = “Fairly confident correct”; 7 = “Completely confident correct”). The computer program recorded the study lists presented along with the responses and confidence levels given at testing. The study lasted 50 - 60 minutes. The participants were given a debriefing form and the researcher responded to their possible queries on the experiment after the study was completed.

5.4. Results

5.4.1. *Memory performance*

A 2(recall type: cued vs. uncued) X 3(number of categories in the study lists: 2-6-24) mixed ANOVA was conducted on retrieval performance (forced-report memory quantity); see Figure 9. The results did not reveal a significant recall type main effect, $F<1$. However, a main effect of the number of categories in the study lists was found significant, $F(2, 116) = 58.721$, $p < .001$, $\eta^2 = .503$. Mean comparisons revealed that retrieval performance for the two-category condition was significantly higher (M = 61.7, SD = 18.9) than that in six-category condition (M = 42.5, SD = 21.5), $t(59) = 7.111$, $p < .001$, which in turn was significantly higher than the performance for the twenty-four-category condition (M = 34.7, SD = 21.6), $t(59) = 3.115$, $p < .001$. The results also showed a two-way interaction between recall type and number of categories in the study lists, $F(2, 116) = 6.640$, $p = .002$, $\eta^2 = .103$.

The retrieval performance reduced as the targets were less and less related to each other both in cued-recall and uncued-recall groups. Whilst the retrieval performance amongst uncued-recall group in twenty-four category, six-category, and two-category target recalls were M = 66.1, SD = 15.0; M = 39.3, SD = .21.6; M = 30.6, SD = 17.0, respectively, the retrieval performance in cued-recall group were M = 57.4, SD= 21.5; M = 45.7, SD = 21.3; M = 38.9, SD = 24.9, respectively. The interaction appeared because the increase in the retrieval performance as a function of increased inter-target association was greater in free-recall group than in cued-recall group; see Figure 9.[18]

5.4.2. *Monitoring performance*

Being a nonparametric measurement, Area Under the Curve (AUC) scores were calculated individually as in Experiment 3 (see section 1.3.4. and Figure 4 for details on the calculation method). Complementing the monitoring performance (AUC), Receiver Operating Characteristics (ROC) curves were also constructed and displayed in Figure 10. Table 13 displays the mean confidence levels and number of targets, cues, and extra-list words, which were reported or withheld (passed) in each group as a function of number of categories in the study lists.

[18] The same ANOVA statistics was conduced on strictly-scored (forced-report) memory quantity. The results indicated exactly the same patterns when liberally scored memory quantity was investigated.

A 2(recall type: cued vs. uncued) X 3(number of categories in the study lists: 2-6-24) mixed ANOVA was conducted on monitoring performance, AUC with recall type as the between-participants factor[19]; see Figure 9. Neither the main effect of recall type, nor the interaction from the ANOVA was significant, largest $F(1, 54) = 2.70, p = .11$. However, the main effect of the number of categories in the study lists was significant, $F(2, 108) = 4.068, p = .02, \eta^2 = .069$. Participants monitoring performance for the twenty-four category list ($M = .94, SE = .01$) was better than for the two-category list ($M = .89, SE = .01$), $t(55) = 3.065, p = .003$, and it was better than for the six-category list ($M = .89, SE = .02$), $t(55) = 2.338, p = .02$. Monitoring performance was comparable between the two-category and six-category lists, $t(55) = .079, p = .94$. As expected, participants monitored their responses better as the number of categories in the study list increased in contrast to memory quantity.[20]

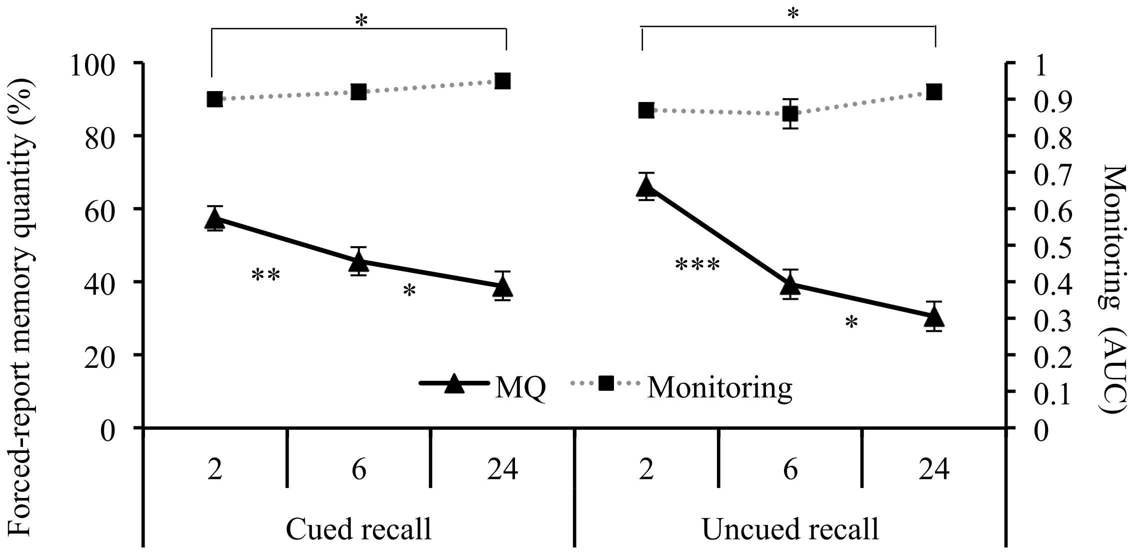


Figure 9. Memory quantity (MQ) and monitoring performance (AUC) as a function of recall type (cued-uncued), and number of categories in study list (2-6-24). Standard errors are shown with the error bars in the figure attached to each mean score.
* = $p < .05$; ** = $p < .01$; *** = $p < .001$.

[19] Four participants were dropped from this analysis, one because of failure to retrieve any targets in one of the experimental conditions (and hence no hit rates), and three additional ones because of extremely low, below-chance monitoring scores (near zero), suggesting that the confidence scale was used opposite to what was instructed (i.e., high confidence was represented by low values on the scale rather than high ones).
[20] The results gathered when strict scoring was adopted showed exactly the same pattern as the results yielded when liberal scoring was investigated.

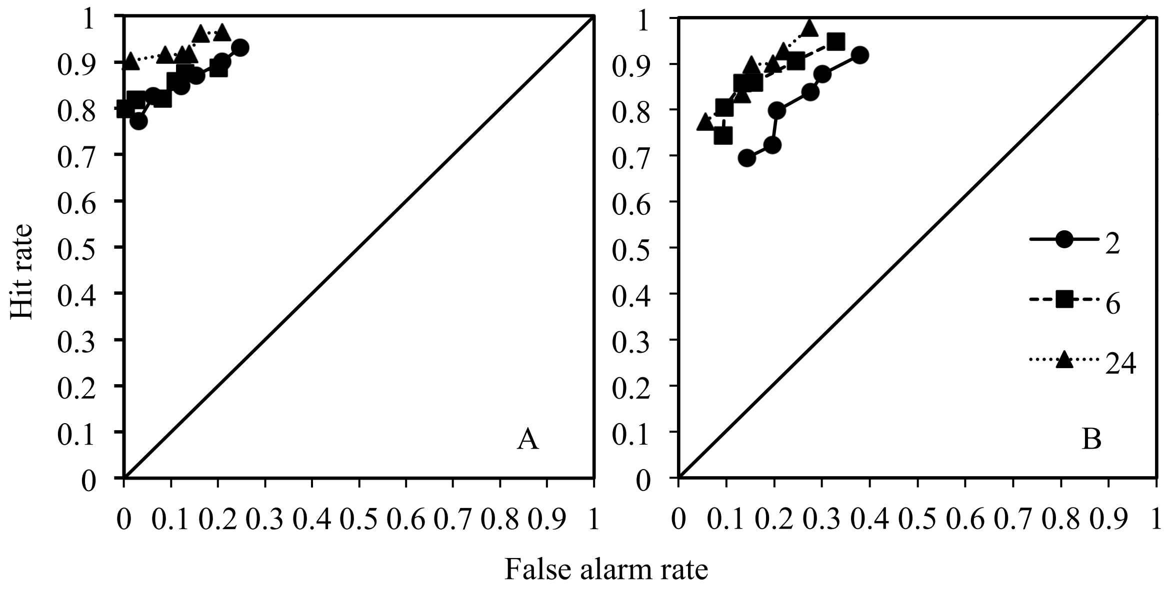


Figure 10. ROC curves constructed as a function of the number of categories in the study list (2-6-24) for the cued-recall group (A) and uncued-recall group (B).

Table 13

Mean Confidence Levels (CF) and Number (Nbr) of Targets, Cues, and Extra-List Words Reported or Passed as a Function of Number of Categories in Study Lists (Two vs. Six vs. Twenty-four) and Group (Cued-Recall Group vs. Uncued-Recall Group)

		Reported								
		Two			Six			Multiple		
Grp		Target	Cue	Extra-list	Target	Cue	Extra-list	Target	Cue	Extra-list
Cr	CF	5.2(1.5)	-	3.0(1.7)	5.7(1.9)	-	2.6(1.3)	5.9(1.8)	-	3.2(1.4)
	Nbr		-	2.2(3.0)		-	3.6(5.6)		-	3.3(5.5)
UCr	CF	5.1(1.8)	3.6(3.1)	3.3(1.6)	5.1(2.1)	3.6(2.0)	3.3(2.0)	5.3(1.9)	4.9(2.1)	3.6(1.7)
	Nbr		0.2(0.6)	2.8(2.7)		0.5(1.3)	3.8(3.9)		0.5(1.1)	4.1(4.6)

		Passed								
		Two			Six			Multiple		
		Target	Cue	Extra-list	Target	Cue	Extra-list	Target	Cue	Extra-list
Cr	CF	2.0(1.2)	-	1.5(1.5)	2.2(1.3)	-	1.7(1.5)	1.7(0.9)	-	1.5(1.5)
	Nbr		-	10.2(5)		-	12.7(5)		-	14.7(6)
UCr	CF	1.8(1.7)	2.3(2.3)	1.7(1.6)	2.9(2.1)	1.7(1.6)	1.6(1.5)	2.5(2.0)	1.9(1.7)	1.3(1.0)
	Nbr		0.6(1.1)	7.5(3.5)		0.7(1.4)	13.9(5)		1.0(1.6)	15.7(4)

Note. Grp= group; Cr= cued recall; UCr = uncued recall. Standard deviations are in parentheses. See Figure 9 to see the total number of targets (% recall), which were reported or passed.

5.5. Discussion

Experiment 4 confirmed the hypothesis that when recall performance for the inter-related items is enhanced as the extent of inter-target association increases (e.g., two-category list target recall), it is accompanied with deterioration of monitoring performance. In other words, enhancing retrieval performance via organizational effects does come with a cost on monitoring performance. Although it has been well documented that organization (e.g., chunking, groupings, or categorization) increases retrieval performance (e.g., Bousfield, 1953; Jenkins, Milk & Russel, 1958; Tulving, 1962; Tulving & Pearlstone, 1966; Cohen, 1966), its effects on metacogitive processes such as on metacognitive monitoring have not been clear. Therefore, a clear dissociation observed in this experiment was important to show that the strategies to enhance retrieval performance (e.g., categorization) might turn out to be a trade-off emerging between memory and metamemory performance.

To be more specific, the results showed that the participants (cued-recall and uncued-recall participants together) had the lowest recall performance in twenty-four category target recall (33%) and the highest one in two-category target recall (53%). This result was inline with the previous findings showing that the number of items recalled increases as the number of categories in the study lists decreases (e.g., Tulving & Pearlstone, 1966; Slamecka, 1968, 1972).

However, the most important finding of the current experiment was observed when memory and monitoring performance of the participants were compared as a function of number of categories in the study lists. It was mainly found that the facilitated generation process, emerging as a result of high inter-target association (such as when study lists had two discrete categories of target items), deteriorated the monitoring performance. When, for instance the ROC curves amongst uncued-recall are further scrutinized (see Figure 10), it seemed that the reason for the outcome in the monitoring performance stemmed from relatively higher false alarm rates (as well as lower hit rates) in the two-category target recall compared to the six-category target recall and even higher false alarm rates (and lower hit rates) in twenty-four category target recall. The depictions on the ROC curves were important to show that participants tended to report some incorrect responses with high confidence when study lists had higher inter-target association. This pattern converges with some earlier investigations on false memory (e.g., Kato, 1985; Roediger & McDermott, 1995; Roediger, Watson, McDermott & Gallo, 2001) that participants mistakenly report incorrect responses as considering that they, indeed, were correct responses.

The calculations on intrusion errors also elucidate the reasons behind the deteriorated monitoring performance when targets were highly inter-related. For instance, let us consider the uncued-recall participants amongst whom a full dissociation between memory and monitoring performance was observed (see Figure 9). As the uncued-recall participants were not provided with any cues at the time of testing, cue type intrusion errors would be naturally expected. However, the percentages of cues reported (reported and withheld together) were found to be quite low in the two-category, six-category, and twenty-four-category target recalls ($M = 2.5$, $SD = 4.46$; $M = 2.78$, $SD = 5.62$; $M = 4.03$, $SD = 6.61$, respectively). Paired-sample t-test results showed that uncued-recall participants did not differ between cues reported in two-category target recall and six-category target recall, $t(29) = -.239$, $p > .05$, between six-category and twenty-four category target recalls, $t(29) = -1.121$, $p > .05$, and between two-category and twenty-four category target recalls, $t(29) = -1.187$, $p > .05$. The errors that were commonly made were the extra-list type of words: the words being neither cues nor studied targets. The percentages of these extra-list words given by the uncued-recall participants were calculated in terms of whether they were reported or withheld. In terms of percentage of 'extra-list words reported', the uncued recall participants had the following scores: $M = 11.8$, $SD = 11.2$ in two-category target recall; $M = 16.0$, $SD = 16.3$ in six-category target recall; and, $M = 17.2$, $SD = 19.3$ in twenty-four-category target recall. Paired-sample t-test results statistically revealed that uncued-recall participants did not differ between extra-list word reported in two-category target recall and six-category target recall, $t(29) = -1.67$, $p > .05$, between six-category and twenty-four category target recalls, $t(29) = -.459$, $p > .05$, and between two-category and twenty-four category target recalls, $t(29) = -1.879$, $p > .05$. Nonetheless, those responses 'withheld' varied substantially and the uncued-recall participants had the highest percentage of incorrect responses passed in twenty-four-category target recall. The percentages of 'extra-list words withheld' were: $M = 19.6$, $SD = 13.7$ in two-category target recall, $M = 42.0$, $SD = 24.2$ in six-category target recall, and $M = 48.2$, $SD = 21.0$ in twenty-four category target recall. The t-test results indicated that uncued-recall participants had lower percentage of extra-list words withheld in two-category target recall compared to six-category target recall, $t(29) = -5.122$, $p < .001$, whereas they did not differ in terms of extra-list words withheld in six-category target recall compared to twenty-four target recall, $t(29) = -1.540$, $p > .05$. Moreover, uncued-recall participants had lower percentage of extra-list type of words withheld in two-category target recall compared to twenty-four category target recall, $t(29) = -7.245$, $p < .001$. In other words, the uncued-recall group could not withhold incorrect responses in two-category target recall as well as, for instance, in twenty-four category target recall. This pattern provides further evidence that when the participants recall targets that are highly related to each other, they could not monitor their correct and incorrect responses as much as they could recall the targets that do not have any (normatively designated) association between each other.

To summarize, Experiment 4 was again found to be converging to the reasoning of Higham & Tam (2005), as well as the findings of Guerin and Miller (2008): Should the generation process of candidates is facilitated via such as organizational effects, recognition of the correct items from amongst the generated candidates becomes difficult. In contrast to Guerin and Miller, however, this experiment directly investigated the effects of high inter-target association on memory and metacognitive functioning via inferring the recognition process in a single retrieval process, recall. In short, Experiment 4 showed that memory enhancement methods such as inter-relatedness amongst to-be-remembered items might come with a cost on monitoring performance.

CHAPTER 6
Experiments 5 and 6

6.1. Introduction

Experiment 4 confirmed that categorical relatedness between to-be-remembered items enhances recall performance of these items. However, the enhancement of retrieval is accompanied by deterioration of monitoring performance. This dissociation between memory and metamemory performance was contingent on the enhancement of the retrieval via a well-known memory enhancement method: increasing the relatedness between to-be-remembered material, such as categorical relationship (e.g., Battig, 1966; Runquist, 1970; Tulving, 1962; Tulving & Pearlstone, 1966; Puff, Murphy & Ferrara, 1971). Hence, confirming the expectations made on the basis of generate-recognize perspective (Bahrick, 1970), Experiment 4 showed that when the generation process of candidate items is facilitated via high interrelatedness between target items, it negatively affects recognition of the targets from amongst the inter-related candidates.

Based on the results gained in Experiment 4, the following prediction was made. If the target materials in paired-associate learning are strongly encoded along with the cues paired with them, then the memory-metamemory dissociation is expected to disappear. It was expected that, rather than utilising the inter-target association to enhance recall performance, the participants would depend more on cues to retrieve the paired target items when those pairs were encoded well enough. In other words, when cue-target pairs are 'individuated', inter-target association would be discarded in favour of a greater reliance on cues to retrieve the target information, since the cues are expected to be more effective memory triggers. As a result, an enhancement in recall performance would not have any catastrophic result on monitoring performance. The current experiment, hence, aimed to see whether the dissociation that was observed in Experiment 4 could be removed by introducing a processing that encourages individuation of the pairs. It was anticipated that the current experiment would show that the existence of inter-target association (e.g., two-category study list) is not always detrimental to monitoring performance, rather it depends on whether such interrelatedness would be used at retrieval or not.

There seemed at least two conventional methods to manipulate encoding level (depth) of cue-target pairs processed in Experiment 5. One method is to manipulate the semantic association between pairs via study lists that include strong vs. weak-associate pairs when keeping the encoding strategy constant between the lists (e.g., both lists are studied with same encoding strategy). The other method is to manipulate the levels of processing at study when keeping the semantic association level between pairs constant between the lists (such as involving only weak-associate pairs). Except the study list used in Experiment 1 that involved high semantic association (SA) in cue-target pairs, the remaining three of the experiments reported hitherto (Experiments 2, 3, and 4) had consistently weak-associate pairs in their study lists[21]. Hence, following the previous studies reported so far, we decided to manipulate the level of processing rather than manipulating the semantic association between the paired stimuli. It should be noted that Experiment 3 manipulated the encoding process of pairs, and this was achieved by pair-focused and target-focused ways of encoding. The results of Experiment 3 showed that the pair-focused encoding strategy yielded (partial) dissociation between memory and monitoring performance only in uncued-recall groups, although the target-focused strategy resulted in the expected dissociation both in the cued-recall (only when liberal scoring was adopted) and in uncued-recall. However, the aim in Experiment 5 was to show whether the full dissociation found in Experiment 4 could be removed with a manipulation on encoding strategy that more strongly differentiates the levels of processing than the manipulation utilised in Experiment 3. Additionally, in contrast to Experiment 3, the participants in the present study all knew in advance of testing that they would be tested on only the target items (the words on the right hand side in the pairs). As a result, the encoding manipulation herein can be considered as a differentiation of the depth of encoding amongst target-focused participants. In this way, interactive imagery was predicted to individuate pairs stronger than the pair-focused way of encoding used in Experiment 3.

The encoding strategies which were instructed to the participants to be used whilst studying/encoding the pairs were: *rote repetition*; simply repeating the material out loud, and *interactive imagery*; subjectively constructing visual images between the paired items. In this manipulation, rote repetition was expected to be ineffective in binding the cue words with their very target words paired with at study. However, interactive imagery was expected to be effective in doing so (e.g., see Dunlosky & Hertzog, 1998a, 1998b). As a result, considering that the cues paired with targets is expected to be effective retrieval aids rather than using the interrelatedness between

[21] Likewise high-SA study lists used in Experiment 1, the study lists in Experiment 6 also had strong semantic association between the cues and the targets paired with them.

target items to guide recall, high inter-target association (such as in two-category target recall) would be discarded as much as possible when pairs are encoded with more relational mnemonic, and so these pairs would be more connected to each other with interactive imagery strategy rather than with rote repetition. In this sense, the aims of the present experiment are twofold: it will test the possibility that the dissociation found clearly in Experiment 4 could be eliminated, and whether the usage of aids in cued-recall retrieval are prioritised. The first retrieval aid seems the utilization of covert cues as much as possible when they are available at test (cf. encoding specificity principle, Thomson & Tulving, 1970). The second one seems the other sources, such as the features and structures of the study lists, i.e., inter-target associations.

As mentioned earlier, it has been well-documented in a considerable number of studies of research manipulating the levels of processing (Craik & Lockhart, 1972) that the deeper the information is processed, the greater is the number of remembered items (e.g., Tresselt & Mayznet, 1960; Shulman, 1971; Craik & Lockhart, 1972; Craik & Tulving, 1975; Shulman, 1971; Stein, 1978; Mantyla, 1986; Koriat & Melkman, 1987; Mantyla & Nilsson, 1988; Hunt & Smith, 1996; Polyn, Norman, & Kahana, 2009b); see section 4.1 in Chapter 4 for an extensive review of the levels of processing perspective. Interactive imagery, for instance, is expected to result in a deeper, long-lasting encoding strategy than such as rote repetition and it has been supported that interactive imagery is a quite effective encoding strategy in retrieval compared to rote repetition (e.g., see Dunlosky & Hertzog, 1998a, 1998b). However, some earlier findings indicated that if the words paired are semantically associated, any technique, including rote repetition, facilitates retrieval (e.g., Dunlosky & Hertzog, 1998a; Hertzog, Price & Dunlosky, 2008). On the other hand, if the word pairs are unrelated -just like the study list used in the current research- an encoding strategy such as using a relational mnemonic like interactive imagery is effective in retrieval (e.g., Dunlosky & Hertzog, 1998b; also see Pavio, 1978). It was, therefore, decided that manipulating encoding strategy (rote repetition vs. interactive imagery) rather than semantic association between pairs (e.g., strong vs. weak) was to be used in this experiment. The main objective was to individuate weakly-associated (unrelated) pairs which are expected to dissolve the (full) dissociation found in Experiment 4.

The study lists had two, six, and twenty-four categories to which target items belong in Experiment 4. However, the decision of which categories, which targets from these categories, and which cue words would be paired with the target items were determined randomly by the computer program in Experiment 4 (see section 5.7. for the details on constructing the study material in Experiment 4). Although following Experiment 4, encoding manipulations (e.g., rote repetition vs. interactive imagery) have been of interest in the current experiment, Experiment 5. Therefore, concreteness and imageability values of the paired-words were needed so as to be used for matching these values between the cues and their paired targets. These values were also needed not to differ between the study lists (such as two vs. twenty-category study lists) to gather clear results not contaminated by such values as much as possible. As a result, the same study lists used in Experiment 4 were not used in Experiment 5.

It was, therefore, expected that the cued-recall participants using interactive imagery would encode the cue-target pairs as individuated as possible, so the retrieval performance in the two-category and the twenty-category target recalls would be comparable. As a result, monitoring performance would also not differ between the two-category and the twenty-category target recalls. In other words, what was expected to be critical was that there would be greater reliance on the cues to remember targets when more relational encoding strategy is used, such as interactive imagery. On the contrary, rote repetition was expected to yield the same dissociation found in Experiment 4 in cued-recall group. The predictions made for the uncued-recall group were, however, different than the ones for the cued-recall group. That is, when participants were not given any explicit cues at testing (uncued), then they were expected to seek a more effective retrieval source, such as the interrelatedness of the targets and/or the cues, if they are remembered.

Although inter-target association turns out to be an effective source to guide retrieval in the cued-recall group, it is expected that the uncued-recall participants will take advantage of interrelatedness of the target items relatively more than the cued-recall participants, based on the findings gained so far. For instance, as it was found in Experiment 2 and Experiment 3, the dissociation is clearer when participants were uncued compared to when they were cued at recall. Therefore, regardless of the level of encoding by which cue-target pairs are studied, the uncued-recall participants were expected to yield the dissociation observed in Experiment 4. Because they were expected to utilise the inter-target association in the retrieval no matter what type of encoding strategy they use. However, the uncued-recall group was expected to have a clearer dissociation when they use rote repetition (e.g., full dissociation) compared to when they use an interactive imagery strategy (e.g., partial dissociation). Because, interactive imagery was expected to let even uncued-recall participants depend on cues more often than rote repetition. The patterns of dissociation for the uncued-recall group in Experiment 3, which manipulated the focus that participants adopt whilst studying the lists, revealed that the uncued-recall group had the same (partial) dissociation in both study types, pair-focused and target-focused. In the present study, however, the uncued-recall group was expected to yield a clearer (full) dissociation when rote repetition is utilised compared to when they use interactive imagery (partial dissociation) by considering that the differentiation of the levels of processes would be greater with the current encoding manipulation (interactive imagery vs. rote repetition) compared to study type manipulation (pair-focused vs. target-focused) (see section 5.2. for the definitions of partial and full dissociations).

6.2. Method

6.2.1. *Participants*

A total of 64 undergraduate and postgraduate students in the University of Southampton, and college students from Totton College, Southampton, participated in the study. The participants spoke English as a first language. They were compensated for their time with either course credits or £5 payment, and were randomly allocated to one of the conditions of the study: cued-recall and uncued-recall groups; see Table 14 for the demographic characteristics of the groups[22].

Table 14

Demographic Characteristics of the Cued and Uncued-Recall Groups in Experiment 5

		Age		Gender	
Recall type	*n/N*	*M*	*SD*	Male	Female
Cued recall	32	24.5	5.63	16 (50.0%)	16 (50.0%)
Uncued recall	32	27.0	5.47	17 (53.1%)	15 (46.9%)
Total	64	25.7	5.65	33 (51.6%)	31 (48.4%)

6.2.2. *Experimental design*

The study used a 2(recall type: cued vs. uncued) X 2(encoding strategy: interactive imagery vs. rote repetition) X 2(number of categories in the study lists: two

[22] Amongst 64 participants, totally 17 participants from Totton College voluntarily took part in the study on an event called ‘Psychology Day’. This event is run annually in collaboration with various schools of psychology in England including the School of Psychology of the University of Southampton and the management of the Totton College in order to build up awareness amongst college students in psychological research.

vs. twenty) X 2(report type: free vs. forced) mixed factorial ANOVA with recall type as the between participants variable, and encoding strategy, number of categories in the study lists, and report type as within participant variables. The dependent variables were forced-report memory quantity (retrieval performance), monitoring performance, and confidence levels.

6.2.3. *Materials*

6.2.4. *Study lists*

Four study lists which composed of 20 word pairs in each (e.g., cue-target) were constructed. Two of the lists were two-category target lists, in which the target words belong to two discrete categories (T1 & T2). In other words, each of the two-category lists, T1 and T2, involved 10 exemplars from two discrete categories. On the other hand, the remaining two lists were multiple-category target lists, in which the target items belong to 20 different categories (M1 & M2). There were two versions of the study lists (T1 & T2; M1 & M2) as the encoding strategy manipulation necessitated: One of the versions of the lists (e.g., T1 and M1) would be studied with interactive imagery and the other version (T2 and M2) would be studied with rote repetition (the order of encoding strategy to be used was counterbalanced; see procedure for details). The categories were selected on the basis of the following criteria: They are mutually exclusive and have at least 10 exemplars available in the norms of Van Overschelde et al. (2004). Also, the targets are all nouns and these targets have comparable free association means within each two-category target list (T1 & T2), which were the percentage of participants in the study of Van Overschelde et al. who gave a particular response to each category name given. As a result of the criteria, the categories of *fruits, animals* (in T1), *pieces of clothing,* and *musical instruments* (in T2) were determined. The target words (10 from each category) were then matched with weakly-associated cue words based on the written frequency of Kucera-Francis (1967) and the concreteness and imageability values of MRC Psycholinguistic Database. Multiple category lists were created based on the obtained values of two-category lists. To be more specific, M1 was matched with T1, and M2 was matched with T2 in terms of the written frequency, concreteness, and imageability values. Since it was a quite painstaking process to construct the lists with respect to matching all of the values at the same time, written frequency and imageability values were determined to be first in the priority than the concreteness values; see Appendix for the study lists used in Experiment 2 (see Appendix E for the study lists used in Experiment 5).

6.2.5. *Procedure*

The participants were tested individually in the cognitive laboratory located in the School of Psychology of the University of Southampton. They were tested in four computerized study-test cycles. The data for Experiments 5 and 6 were collected at the same time. Hence, the study had two main parts of data collection, one for Experiment 5 and one for Experiment 6. In the first part, all of the participants -regardless of the group they were in (cued or uncued)- started with the practice test and continued with the actual study of Experiment 6, which tested the recall performance of the participants after they studied a single study list. In the second part, the participants were informed that they would be starting another part of the study (Experiment 5), which had nothing to do with the first part (Experiment 6). Then, they continued to Experiment 5 by starting with the practice test and then go on with the actual study phase of it.[23]

Participants studied two two-category lists (list-1 and list-2) and two twenty-category lists (list-1 and list-2). The encoding strategy variable (interactive imagery vs. rote repetition) was crossed across the number of categories in the study list variable (two vs. twenty). That is, each type of encoding strategy was used for each study lists having two different numbers of categories in the study lists. The order of study list presentation and the order of encoding strategy used were counterbalanced. However, presentation orders, which dictated a successive shift from one strategy to another, were discarded. It was thought that those combinations of encoding strategy counterbalancing that necessitated a consecutive change from one strategy to another between lists might have created a difficulty for the participants. As a result, there appeared eight different study versions to which the participants were randomly allocated. In order to let the participants understand how the encoding strategies would be performed, the participants took a practice study, in which their responses were not recorded. In the practice phase, the participants studied five cue-target pairs (*advice-negative*, *law-proud*, *fifty-quiet*, *idea-round*, and, *angle-permit*) with interactive imagery

[23] Since all of the participants always had the study-test cycle of Experiment 6 in the first part, there was no retroactive interference affecting the Experiment 6. The first part (Experiment 6) was controlled for all of the participants who passed to Experiment 5. In the second part, the order of study-test cycles were counterbalanced in terms of number of categories (two vs. twenty) and encoding strategy (interactive imagery vs. rote repetition), hence, there was again no high chance of proactive interference caused by Experiment 6 because of counterbalancing of the study-test cycles, except fatigue towards the last study-test cycles; see Figure 19 for the experimental procedure for collecting data for Experiments 5 and 6, and Table 22 for the versions of the study-test orders in Appendix E.

and they recalled the target words (the words on the right hand side in the pairs) from the just-presented list. They were asked to construct a visual image or a scene that involved the words paired within the presentation time of each of the pairs (4-sec). For instance, if the word pair is 'KITCHEN – SKY', they could imagine a scene such as 'imagining yourself in the kitchen looking from the window to the sky'. This practice was followed by the recall of the five target words from the just presented list. After practicing with interactive imagery, the participants were asked to study another 5 word pairs using the rote repetition encoding strategy. They studied the word pairs (*model-famous*, *budget-session*, *total-energy*, *character-demand*, and *pale-strike*) during which they repeated each word pair out loud for 4-sec. An optimal presentation rate that should have let the participants construct images, but at the same time, which should not have allowed them to shift from rote repetition to imagery when they were repeating the word pairs out loud, was needed. Therefore, the slowest possible presentation rate for using interactive imagery strategy properly was needed, which was decided to be 4 seconds in this experiment.[24] Whether the cue words were provided to the participant at the time of testing or not depended on the group to which the participant was allocated (cued-recall vs. uncued-recall). Once the participants finished the practice tests, they were warned that they were about to start the actual study in which their responses would be counted.

As in Experiments 3 and 4, the participants were presented with a study list of 20 word pairs. They encoded/studied the word pairs either with an interactive imagery or rote repetition strategy, depending on the order of study version to which they were allocated. All of the participants were instructed at each study phase that they would be responsible for remembering and reporting the target words, which were the words on the right hand side of the pairs. The participants initiated the list presentations themselves by clicking on a 'start the presentation' button located on the computer screen when they felt ready. The word pairs were presented with capital letters and

[24] For instance, Dunlosky and his colleagues (e.g., see Dunlosky & Hertzog, 1998b), for instance, manipulated the presentation times (4 sec. vs. 8 sec.) when the study materials were studied with two different encoding strategies: rote repetition or interactive imagery. The results of Dunlosky et al. (1998b) showed that regardless of the presentation time, interactive imagery was more effective on memory performance compared to rote repetition. Additionally, it was suggested by John Dunlosky (J. Dunlosky, personal written communication, 18th May 2010) that any presentation time between 4-10 seconds is equally effective for constructing a visual image between pairs. Therefore, the word pairs in Experiment 5 were presented with a rate of 4 seconds with 1 sec inter-stimulus interval.

separated with a hyphen between words (e.g., PALACE – SISTER). After the presentation of each list, they solved some moderate difficulty algebra calculations or Sudoku puzzles for 5 minutes as a filler activity. The computerized-testing phases, during which the responses were recoded electronically, were exactly the same as they were in Experiment 4 (see section 5.8. of Chapter 5 for the details). Unlike previous experiments, each word pair appeared on the computer screen for 4 seconds with 1-second inter-stimulus interval. After the study was completed, the participants were given a written debriefing statement (see Appendix E) and the experimenter responded to their queries. Being self-paced during the testing phase, the study lasted between 55 and 65 minutes.

6.3. Results

6.3.1. *Memory performance*

In order to investigate actual retrieval performance, a 2(recall type: cued vs. uncued) X 2(number of categories in the study lists: two vs. twenty) X 2(encoding strategy: interactive imagery vs. rote repetition) mixed ANOVA was conducted on forced-report memory quantity; see Figure 11. The results showed that recall type main effect was not significant, $F<1$. However, the number of categories main effect was significant, $F(1, 62) = 126.079$, $p< .001$, $\eta^2 = .670$: Retrieval performance was significantly higher in two-category target recall than in twenty-category target recall ($M = 62.5$, $SE = 2.0$; $M = 37.4$, $SE = 2.0$, respectively). Recall type and number of categories had an interaction effect on retrieval performance, $F(1, 62) = 34.730$, $p< .001$, $\eta^2 = .359$. Pair-wise mean comparisons indicated that the difference between retrieval performance in two-category and twenty-category recall was greater amongst the uncued-recall group ($M = 67.5$, $SD = 16.1$; $M = 29.1$, $SD = 16.0$, respectively), $t(31) = 12.845$, $p< .001$, than the difference observed in the cued-recall group ($M = 57.6$, $SD = 17.2$; $M = 45.6$, $SD = 17.4$, respectively), $t(31) = 3.574$, $p = .001$. The results also revealed that the main effect of encoding strategy was significant, $F(1, 62) = 56.014$, $p< .001$, $\eta^2 = .475$: The participants had higher retrieval performance when they studied the word pairs with interactive imagery ($M = 59.1$, $SE = 2.4$) than when they employed rote repetition ($M = 40.8$, $SE = 1.8$). Encoding strategy and recall type showed a significant two-way interaction, $F(1, 62) = 37.390$, $p< .001$, $\eta^2 = .376$. Pair-wise mean comparisons showed that whereas cued-recall group had higher retrieval when they used interactive imagery ($M = 68.3$, $SD = 21.3$) compared to when they used rote repetition ($M = 34.9$, $SD = 16.9$), $t(31) = 7.456$, $p< .001$, uncued-recall group did not differ in terms of retrieval between when they used interactive imagery (*M* = 50.0, *SD* = 17.2) and when they used rote repetition (*M* = 46.6, *SD* = 11.9), $t(31) = 1.669$, $p > .05$. In other words, when participants were cued to recall at testing, encoding cue-target pairs via interactive imagery helps in retrieving the targets greater than rote-repetition could. On the other hand, when cues encoded with targets via interactive imagery were not represented at testing (uncued recall), there seemed no additional benefit of interactive imagery over rote repetition on retrieval (most probably due to the taking advantage of ITA in uncued recall).

The results also indicated a significant interaction effect between number of categories and encoding strategy, $F(1, 62) = 11.162$, $p < .001$, $\eta^2 = .153$. Pair-wise mean comparisons showed that the participants had greater retrieval performance in two-category target recall compared to twenty-category target recall when they adopted rote repetition encoding strategy (*M* = 55.9, *SD* = 23.0; *M* = 25.6, *SD* = 16.5), $t(63) = 9.685$, $p < .001$, and this difference was observed greater when rote repetition was used compared to when interactive-imagery was used (*M* = 69.1, *SD* = 21.0; *M* = 49.1, *SD* = 28.3, respectively), $t(63) = 6.204$, $p < .001$. Lastly, there was no interaction between recall type, number of categories in the study lists, and encoding strategy, $F<1$.[25]

6.3.2. *Monitoring performance*

In order to investigate monitoring performance of the participants, Area Under the Curve (AUC) scores were measured just like in the previous experiments.

First, a 2(recall type: cued vs. uncued) X 2(number of categories in the study lists: two vs. twenty) X 2(encoding strategy: interactive imagery vs. rote repetition) mixed ANOVA was conducted on AUC scores[26]; see Figure 11. The results revealed

[25] The results gathered when strictly scored forced-report memory quantity revealed exactly the same patterns when liberally scored forced-report memory quantity was investigated.

[26] The results gathered when strictly scored monitoring performance yielded exactly the same pattern as did liberally-scored results, except that the number of categories and recall type showed a significant interaction effect, $F(1, 22) = 6.089$, $p = .022$, $\eta^2 = .22$. The interaction appeared because cued-recall participants monitored better in twenty-category target recall than in two-category target recall (*M* = .91, *SD* = .12; *M* = .70, *SD* = .10, respectively), $t(29) = 6.520$, $p < .001$, whereas uncued-recall participants monitored comparably well in twenty-category target recall and in two-category target recall (*M* = .78, *SD* = .18; *M* = .57, *SD* = .28, respectively), $t(9) = 2.162$, $p = .06$.

that recall type and encoding strategy main effects as well as the interaction effect between encoding strategy and recall type were not significant, all *F*s<1. However, the main effect of number of categories in the study lists was significant, $F(1, 54) = 131.824$, $p < .001$, $\eta^2 = .71$. Participants had significantly poorer monitoring performance in two-category target recall ($M = .70$, $SE = .03$) than in twenty-category target recall ($M = .90$, $SE = .02$). Lastly, the results did not show any significant interaction effects between number of categories and recall type, encoding strategy and number of categories, and between encoding strategy, number of categories, and recall type, all *F*s<1.

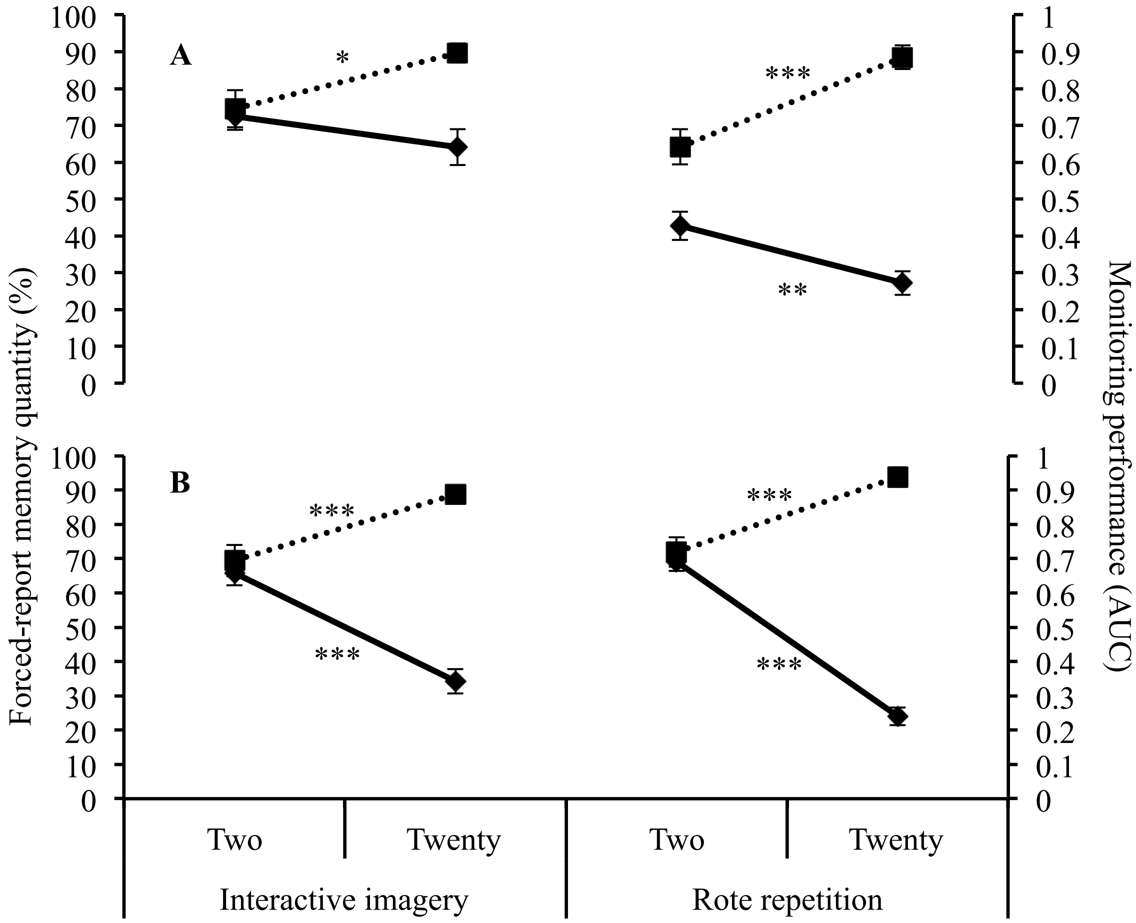


Figure 11. Forced-report memory quantity and monitoring performance (AUC) of the cued recall-participants (A) and of the uncued-recall participants (B) as a function of number of categories in the study lists (two vs. twenty) and encoding strategy (interactive imagery vs. rote repetition). Straight lines (—) display forced-report memory quantity, dashed lines (······) display monitoring performance. Error means are shown with error bars attached to each mean score.

* = $p < .05$; ** = $p < .01$; *** = $p < .001$.

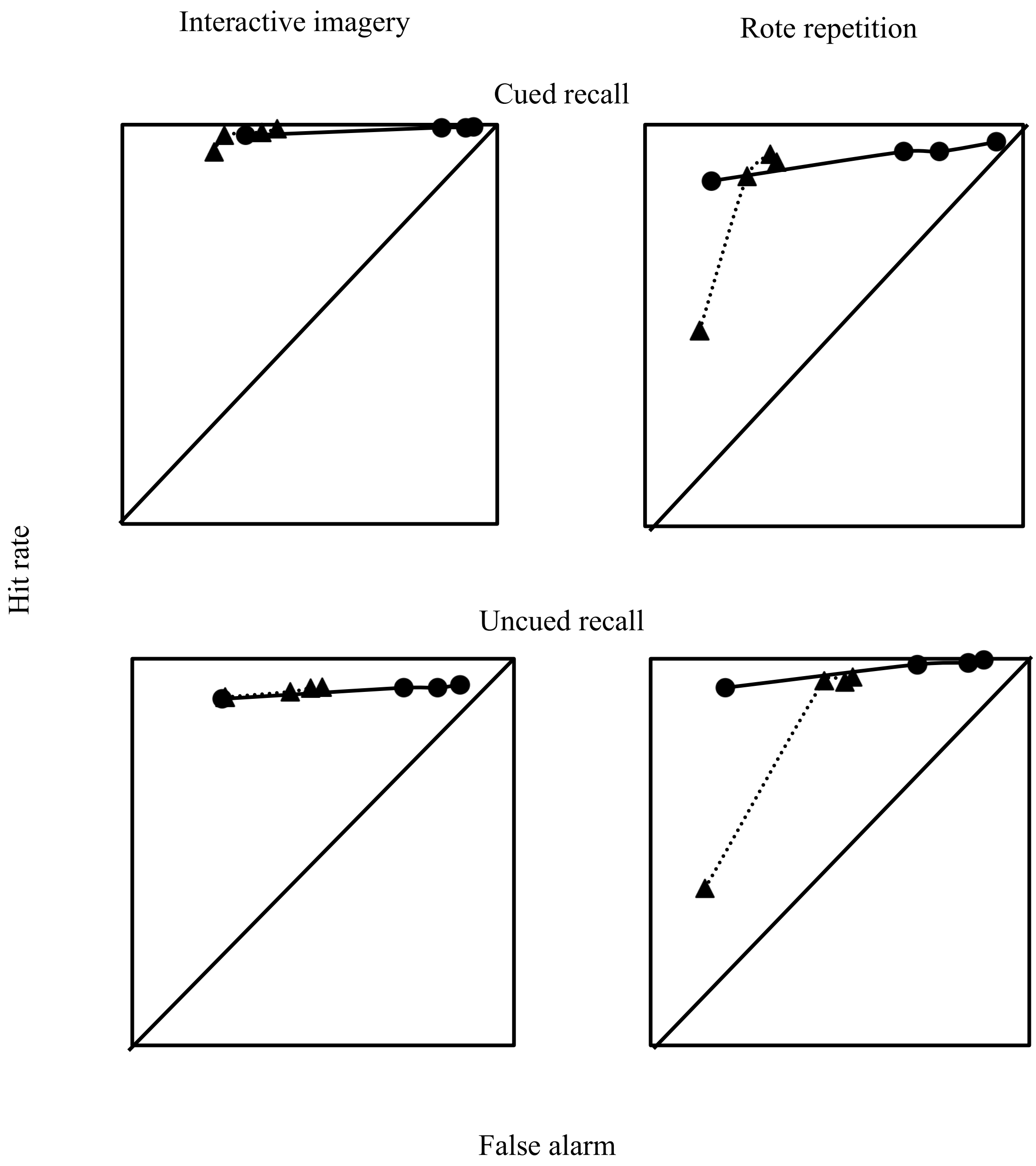


Figure 12. ROC curves displaying monitoring performance (AUC) in cued-recall group and uncued-recall groups as a function of study type (interactive imagery vs. rote repetition) and number of categories in the study lists (two vs. twenty). The two figures at top display the performance of the cued-recall group, whilst the two figures at bottom display the performance of the uncued-recall group. Dashed lines show monitoring performance in two-category target recalls and straight lines show monitoring performance in twenty-category target recalls. It should be noted that the numbers of the cases having confidence levels of 5, 6, or 7 were quite low in both cued and uncued-recall groups. Hence, the ROC curves were displayed at 1+, 2+, 3+, and 4+ cumulative confidence levels.

Table 15

Mean Confidence Levels (CF) and Number (Nbr) of Targets, Cues, and Extra-List Words Reported and Passed as a Function of Study Type (Interactive Imagery vs. Rote Repetition), Number of Categories in Study Lists (Two vs. Multiple) and Group Cued-Recall Group vs. Uncued-Recall Group)

		Reported					
		Interactive imagery					
		Two			Multiple		
		Target	Cue	Exta-list	Target	Cue	Exta-list
Cr	CF	4.6 (1.9)	-	3.2 (2.5)	6.6 (0.6)	-	5.6 (1.6)
	Nbr		-	0.4 (0.8)		-	1.2 (2.2)
UCr	CF	5.3 (1.8)	4.1 (3.0)	4.7 (2.1)	6.1 (1.4)	6.1 (1.5)	4.2 (1.8)
	Nbr		0.1 (0.4)	0.8 (1.2)		0.4 (0.5)	1.0 (1.7)
		Rote repetition					
Cr	CF	4.8 (1.7)	-	4.3 (2.3)	5.9 (1.4)	-	3.5 (1.9)
	Nbr		-	0.9 (1.5)		-	1.0 (1.6)
UCr	CF	5.4 (1.5)	3.7 (2.9)	4.7 (2.0)	6.4 (1.2)	5.0 (1.5)	3.3 (1.7)
	Nbr		0.0 (0.1)	0.6 (0.8)		0.3 (0.5)	1.2 (1.8)
		Withheld					
		Interactive imagery					
		Two			Multiple		
		Target	Cue	Exta-list	Target	Cue	Exta-list
Cr	CF	1.7 (0.9)	-	2.0 (1.6)	4.5 (3.0)	-	1.1 (0.5)
	Nbr		-	2.3 (2.1)		-	2.2 (2.0)
UCr	CF	2.9 (2.2)	3.2 (2.5)	2.8 (2.2)	3.0 (2.2)	1.4 (0.8)	1.4 (1.1)
	Nbr		0.2 (0.4)	2.4 (1.8)		0.3 (0.4)	4.8 (2.6)
		Rote repetition					
Cr	CF	3.1 (2.4)	-	3.0 (2.1)	2.9 (1.7)	-	1.1 (0.2)
	Nbr		-	4.8 (2.3)		-	5.8 (2.5)
UCr	CF	2.6 (2.1)	2.0 (1.4)	2.6 (1.9)	2.8 (2.5)	1.2 (0.4)	1.2 (0.4)
	Nbr		0.1 (0.2)	2.3 (1.9)		0.2 (0.3)	5.9 (2.3)

Note. Cr = cued-recall group; UCr = uncued-recall group. Standard deviations are in parentheses. See Figure 11 to see the total number (%) of targets reported and passed.

6.4. Discussion

The results confirmed that the dissociation between memory and monitoring as a function of the inter-relatedness between to-be-remembered items emerges when participants are put in the conditions in which the inter-target association is utilised to guide recall. The results showed once more that inter-target association facilitated the retrieval performance, however, it did not have the same merit on metacognitive monitoring performance. Counter to the expectation, however, the dissociation did not diminish when such utilization is weakened by having participants depend more on individuated pairs instead of taking advantage of inter-target association (via interactive imagery). This result was, however, found when recall type was collapsed acrossed. When cued and unuced recall groups were considered separately, the following patterns were observed.

As expected, the dissociation between memory and monitoring performance was observed clearer in cued-recall group when they employed rote repetition encoding strategy rather than when they adopted interactive imagery (see Figure 11): Individuation of the pairs established with interactive imagery resulted in a less dependency on categorical relatedness amongst target items. Although it did not disappear completely, the dissociation could be attenuated in cued recall group when the pairs were individuated. The dissociation was, however, clearly observed when rote-repetition was used since it was expected that the cued-recall participants would depend more on inter-target association when re-presented cues turns out to be ineffective in retrieving targets due to rote repetition encoding strategy. This was important to reveal that the cued-recall participants seemed to take advantage of some other sources available, such as the categorical relatedness between targets (ITA), when the cues are not effective in retrieving the stored information. In other words, the utilization of inter-relatedness between targets facilitates the generation process; However, it result in less possibility to recognize the targets from amongst these generated candidates (indexed with poorer monitoring performance of the cued-recall participants that used rote repetition in two-category target recall than in twenty-category target recall).

Unlike the cued-recall group, the dissociation between memory and metamemory performance was found in uncued-recall regardless of the encoding strategy, interactive imagery or rote repetition. However, it was observed that the dissociation was much clearer when the uncued-recall participants used the rote-repetition strategy (full dissociation) compared to when they used interactive imagery (partial dissociation). It is herein noteworthy that although interactive imagery results in higher recall performance than rote repetition (e.g., see Dunlosky & Hertzog, 1998a, 1998b; Hertzog, Price & Dunlosky, 2008), the uncued-recall participants had comparable memory performance in two-category target recall regardless of the encoding strategy they used. That is, the pairs studied with rote repetition yielded comparable retrieval performance in the uncued-recall group (M = 68.5, SD = 16.1) to the retrieval performance when interactive imagery was utilised (M = 66.8, SD = 20.8), *t*(31) = .158, *p*> .05. This finding was attributed to the fact that the uncued-recall participants seemed to utilise inter-target association to aid retrieval both when they used interactive imagery and when they used rote-repetition. Otherwise, the retrieval performance when interactive imagery was used should have been higher than the retrieval when rote repetition was used, only if the cues were depended more in remembering rather than relying on the categorical relatedness between targets. Experiment 3 also has converging evidence to this implication in the way that the uncued-recall participants seem to seek any available source (in this case, ITA) to retrieve the target items rather than trying to remember the cues to help retrieval of the targets with them.

However, the uncued-recall group showed higher memory quantity in twenty-category target recall when they used interactive imagery (*M* = 34.2; *SD* = 20.6) compared to when they used rote repetition (*M* = 24.6, *SD* = 18.4), *t*(31) = 3.604, *p* = .001. In other words, when targets were not related the uncued-recall participants seemed to utilise cues that aided retrieval of the targets since the deeper encoding strategy (interactive imagery) seemed to be the only possible factor to yield this difference. Therefore, the occurrence of the (partial) dissociation amongst the uncued-recall when they used interactive imagery was attributed to the fact that the uncued-recall participants took advantage of ‘cues’ (remembered with more relational encoding, interactive imagery) for unrelated targets (twenty-category) so that they monitored their responses as good as when they monitored their responses for related-targets (two-category). In short, the uncued-recall participants seemed to utilise different aids to retrieve the target items: the inter-target association when targets were related and the retrieval of cues when targets were unrelated, which was an expected finding.

Moreover, considering both of the findings that the cued-recall group did not seem to use target-association as much as the uncued-recall group and the uncued-recall participants relied on target associations in the study lists when such associations existed (MQ-str was higher for two-category target recall than twenty-category target recall in uncued-recall group), the following inference was drawn. The participants seem to depend on sources of retrieval in a certain order. It seemed that the reinstated cues are utilised as a first source in retrieval (when they are available), and then other available sources such as target-associations are used if cues are not reinstated overtly. This implication converges to the findings of encoding specificity principle (Thomson & Tulving, 1970) and suggests that reinstated cues are quite effective retrieval aids when available and strongly encoded with the t-b-r items.

To summarise, the present study found that the dissociation between memory performance and monitoring that emerged as a result of the high inter-target association disappeared via using a more relational encoding strategy in cued-recall. This finding thereby confirms that the dissociation emerges as long as the participants take advantage of categorical relationship to guide their recall instead of depending more on the cues. It should also be noted that the knowledge gathered about the list structure (in this case, the existence of ITA) turns out to be a more strong source to guide retrieval of the target material (such as generating the candidate items) when participants are not provided with any cues at the time of testing compared to when they are prompted with cues at testing, which is parallel to the observations that categorisation is more clearly observed in uncued recall than in cued recall or paired-associate learning experiments (i.e., Tulving, 1962; Tulving & Pearlstone, 1966; Puff, Murphy & Ferrara, 1971; see Battig, 1966; Runquist, 1970 for exception).

6.5. EXPERIMENT 6

6.6. Introduction

Experiment 4 found a clearer dissociation between memory and monitoring performance as a function of ITA level and it raised the following question: Does the dissociation appear because the participants utilise the category relatedness between targets strategically or do they merely retrieve some targets and these targets automatically facilitate the activation of other related targets? In other words, does the dissociation appear as a result of a strategic or an autonomic processing?

In order to respond to the above question, a new experimental design was constructed. The current study manipulated the contextual meaning of the target materials in a way that the meaning of the targets would either construct a particular category with other target items or not. In order to do so, the target items were homographs, which are words written exactly the same but have different meanings depending on the context in which they are used. For instance, as reported in The New Reading Teacher's Book of Lists (Fry, Foundtoukids, & Polk, 1985), the verb 'impress' has at least two meanings: 'have a strong effect on' and 'take by force', and its meaning changes as the context in which it is used varies. In the present study, hence, the cue words were manipulated so that the targets' meaning (the to-be-remembered items) changed due to the change in context. This resulted in two study lists. When the target items were related to each other in terms of a categorical association such as 'body parts' category, the word pairs (cue-target) were, for instance, 'sight-eye', 'lip-mouth', 'pain-chest' etc. In this study list, the target words (eye, mouth, and chest) shared the common meaning that they are all an exemplar of the body parts category. This study list referred to category-implied (CI) list. On the other hand, no-category-implied (NCI) study list had the cue-target pairs, such as 'needle-eye', 'river-mouth', 'drawer-chest' etc. As a result, the same target items in the category-implied list were studied in different semantic contexts (e.g., eye of a needle, mouth of a river, or chest of drawers) and these targets did not share a close, common meaning with each other as the same targets in CI list.

The most important distinction between the current study and the previous ones is the level of the semantic association between cues and targets: whereas the semantic association in cue-target pairs was consistently weak in the previous experiments reported (except high-SA word pairs in Experiment 1), the cue words were strong-associates of the targets in the current experiment. It had been inferred from the results of the previous experiments (e.g., Experiments 4 and 5) that the participants realised the categorical relationships and constructed higher order units to guide retrieval of the inter-related targets, particularly when the number of categories in the list is reduced (when the category size was higher). In the previous experiments reported so far, the contexts of the targets in which the target items were studied (e.g., the meanings of the targets that were implied by the cues) had been irrelevant so as to construct particular category between targets. It was only the high inter-target association between targets, which could, arguably, let participants construct higher order units. Hence, in the current experiment, the following argument was tested. It was indeed a strategic processing that let construction of a categorical relationship, but not solely an automatic process resulting in elicitation of one (or several) targets to activate the retrieval of other related targets, which could have facilitated the generation process that thereby deteriorated the recognition process of the correct target from amongst the candidates, particularly when the lists had a small number of categories (e.g., two).

It was, therefore, hypothesized that the cued-recall participants who study CI lists would have higher memory quantity compared to the cued-recall participants who studied NCI lists. Hence, dissociation between memory and monitoring performance was expected in the cued-recall group. Still adopting the same expectation held for the uncued-recall group in Experiment 5, it was thought that the uncued-recall group would seek further help at the time of testing since they are not provided with any explicit cues at testing, unlike the cued-recall group. Therefore, the uncued-recall group would utilise the inter-target association to guide and enhance their retrieval performance when the associations were encoded at study (CI list) so that dissociation between memory and monitoring performance was also expected in the uncued-recall group as well.

6.7. Experimental design

The experiment used a 2(recall type: cued vs. uncued) X 2(implied categories: yes vs. no) X 2(report type: free vs. forced) mixed design with recall type and implied categories as between participants variables, and with report type as within participants variable. As in Experiments 3, 4, and 5, the dependent variables were memory quantity, monitoring performance, and confidence levels.

6.8. *Method*

6.8.1. *Participants*

The data for Experiments 5 and 6 were collected with a single experimental procedure. This procedure had two successive parts. The first part involved a study-test cycle for Experiment 6 and the second part involved four study-test cycles for Experiment 5 (see footnote 22 of section 6.7. for details). Therefore, all of the participants (*N* = 64) in Experiment 5 participated in Experiment 6 as well. However, half of the cued-recall group (*n* =16) and half of the uncued-recall groups (*n* =16) in Experiment 5 were randomly allocated to the conditions of Experiment 6, in which study lists had category-implied homographs (CI list). The remaining half (*n* = 16 in each) were again allocated to the conditions randomly where study lists had no-category-implied homographs (NCI list). In other words, the participants were randomly allocated to one of the four groups: cued-recall category-implied, cued-recall no-category-implied, uncued-recall category-implied, and uncued-recall no-category implied groups. Table 16 displays the demographic characteristics of the groups in Experiment 6.

Table 16

Demographic Characteristics of The Cued and Uncued-Recall Groups in Experiment 6

			Age		Gender	
Recall Type	Whether targets construct categories	*n/N*	*M*	*SD*	Male	Female
Cued	Yes	16	24.8	6.0	7 (44%)	16 (56%)
	No	16	24.1	5.3	9 (56%)	7 (44%)
Uncued	Yes	16	26.0	3.5	9 (56%)	15 (47%)
	No	16	28.0	6.9	8 (50%)	8 (50%)
	Total	64	25.7	5.7	33(52%)	31 (48%)

6.8.2. *Materials*

6.8.3. *Study lists*

Two study lists involving 20 word pairs (e.g., cue-target) were constructed using the same 20 target words in each. However, these targets were paired with 20 cue words which primed the meaning of the targets either to construct particular categories between target words or not (CI and NCI lists, respectively). The criteria to define the categories obeyed the following rules. The categories were mutually exclusive and had at least 10 dominant exemplars available in the category norms of Van Overschelde et al. (2004). More importantly, however, the targets must be homographs. As a result, the target words were 10 exemplars of two categories: body parts (*arms, chest, ear, eye, hand, head, leg, mouth, neck,* and *tongue*) and colours (*black, brown, blue, pink, grey, green, orange, red, gold,* and *white*). The mean free association norms to the terms ‘body parts’ (*M* = 47.8, *SD* = 23.4) and ‘colours’ (*M* = 63.1, *SD* = 32.1) were fairly high. The determination of the cues was as follows. For CI study list, half of the cue words implied the meaning of a ‘body part’ of the targets and half of them implied the meaning of a type of ‘colour’. For NCI, however, the cue words were different from the ones in CI study list. Therefore, the cues in NCI list implied any meanings of the targets, except the meanings of a ‘body part’ or a type of ‘colour’ as in CI list. As a result, the CI study lists had the following cue-target pairs: *hug-arms, pain-chest, listen-ear, sight-eye, wash-hand, hat-head, run-leg, food-mouth, broken-neck, lick-tongue*, *hair-brown, ocean-blue, dress-pink, cloud-grey, grass-green, sun-orange, rose-red, leaf-gold, snow-white*. The target words in the first 10 pairs invoke a body part meaning and the remaining 10 pairs meant a type of colour. On the other hand, NCI study list had the following word pairs: *weapon-arms, drawer-chest, corn-ear, needle-eye, help-hand, tail-head, table-leg, river-mouth, bottle-neck, language-tongue, evil-black, sad-blue, gordon-brown, elephant-pink, old-gray, amateur-green, tree-orange, communist-red, metal-gold,* and *british-white.* The meanings of the target words were checked on an English Dictionary and the pairs constructed were crosschecked with a native English speaker colleague; see Appendix F for piloting of study words.

6.8.4. *Procedure*

The participants were tested individually in a quiet and dimly lit cognitive laboratory located in the School of Psychology of the University of Southampton. They were tested in ‘a single computerized study-test cycle’. Runtime Revolution software computer program was used for the study and test phases of the experiment. All of the participants were instructed that they would be presented with a list of paired words and be responsible for remembering the target words (right hand side words in the pairs) from the just presented list. The presentation of the word pairs and the testing phase were exactly the same as it were in Experiments 4 and 5 (see e.g., section 5.3.5. in Chapter 5). The pairs were presented for duration of 3-sec with 1-sec inter-stimulus interval.

The participants were given a practice test, in which they studied five word pairs (e.g., cue-target) and were tested on recalling the target words in these pairs. The practice items (*phone-birth*, *maths-box, paper-sheep*, *wall-fanatic*, *hire-angry*) were different than the ones used in the rest of the study. After the presentation of the study lists, the participants had a filler task lasting five minutes and during which they either solved some algebraic calculations or Sudoku depending on their preferences. They were tested the same way as participant in Experiments 4 and 5. The participants were debriefed with a single written debriefing form after the second phase of the study (Experiment 5) was completed (see Appendix E).

6.9. Results

6.9.1. *Memory and monitoring performance*

In order to investigate memory performance, forced-report memory quantity was calculated for each participant. The monitoring performance of the participants was again calculated with Area Under the Curve (AUC) scores, just like they were measured in the previous experiments. The mean scores of memory as well as monitoring performance of each group as a function of category implication (yes-no) are displayed in Figure 13. Receiver Operating Characteristics (ROC) curves are displayed in Figure 14, and mean numbers and confidence levels of targets, cues, and extra-list words reported and withheld are displayed in Table 17.

A 2(recall type: cued vs. free) X 2(category implication: yes-no) ANOVA was conducted on forced-report memory quantity; see Figure 13. The results showed that main effect of recall type was significant, $F(1, 60) = 20.944$, $p< .001$, $\eta^2 = .259$. The cued-recall participants had greater retrieval performance (M = 81.6, SE = 3.2) than the uncued-recall participants (M = 60.6, SE = 3.2). However, the main effect of category implication was not significant, $F(1, 60) = 2.468$, $p>.05$. The results indicated a two-way interaction effect between the variables of recall type and category implication, $F(1, 60) = 8.627$, $p = .005$, $\eta^2 = .126$. Pair-wise mean comparisons showed that whereas cued-recall group had better retrieval performance in CI list target recall compared to

NCI list target recall (*M* = 91.9, *SD* = 10.2; *M* = 71.3, *SD* = 18.8, respectively), $t(30) = 3.884$, $p = .001$, uncued-recall group had comparable retrieval performance in CI list target recall and NCI list target recall (*M* = 57.5, *SD* = 23.2; *M* = 63.8, *SD* = 18.7, respectively), $t(30) = -.839$, $p > .05$.

In order to investigate monitoring performance, a 2(recall type: cued vs. uncued) X 2(category implication: yes-no) ANOVA was conducted on Area Under the Curve (AUC) scores; see Figure 13. The results showed that recall type main effect and interaction effect between recall type and category implication were not significant, $F(1, 52) = 1.826$, $p > .05$, $F < 1$, respectively. Main effect of category implication was, however, marginally significant, $F(1, 52) = 3.926$, $p = .053$, $\eta^2 = .07$: Participants monitored their responses in the recall of category-implied targets better than no-category implied targets (*M* = .91, *SE* = .03; *M* = .82, *SE* = .03).[27]

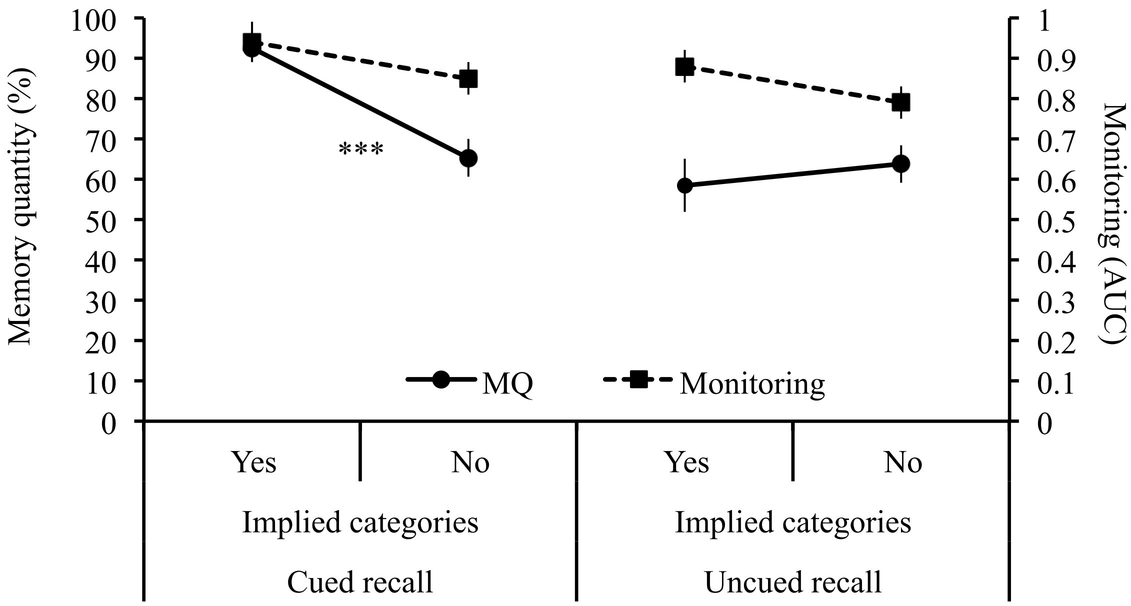


Figure 13. Forced-report memory quantity (MQ) and monitoring performance (AUC) of the cued-recall group and of the uncued-recall group as a function of whether targets construct categories (yes-no). Error means are shown with error bars attached to each mean score.

* = $p < .05$; *** = $p < .001$.

[27] The same ANOVA statistics conduced on strictly scored forced-report memory quantity as well as on strictly scored monitoring performance. These results revealed exactly the same patters when liberally scored performance was investigated, except recall type main effect on monitoring performance was significant, $F(1, 60) = 4.635$, $p = .035$, $\eta^2 = .072$: The uncued-recall participants showed better monitoring performance (*M* = .68, *SE* = .05) than the cued-recall participants (*M* = .83, *SE* = .05).

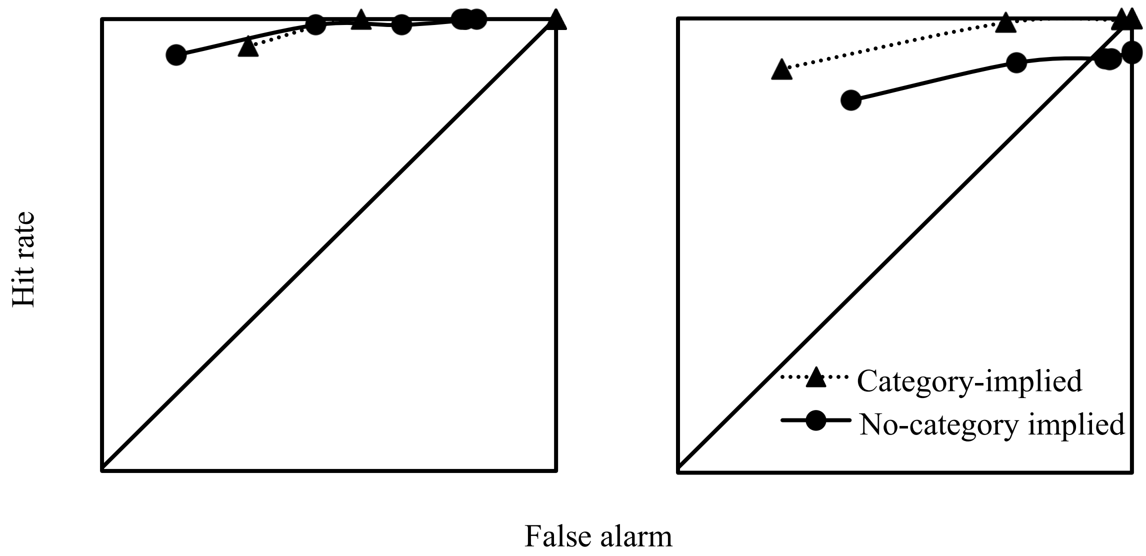


Figure 14. ROC curves displaying response-contingent hit and false alarm rates in cued-recall (A) and uncued recall groups (B) as a function of implied categories (category implied vs. no-category implied).

Table 17

Mean Confidence Levels (CF) and Number (Nbr) of Targets, Cues, and Extra-List Words Reported and Withheld as a Function of Whether Study List Implied Categories or Not (Category-Implied vs. No-Category Implied) in Cued and Uncued-Recall Groups

		Reported			Withheld		
		Target	Cue	Extra-list	Target	Cue	Extra-list
Cued recall							
CI	CF	6.4 (0.9)	-	3.1 (1.2)	1.3 (0.4)	-	1.2 (0.2)
	Nbr		-	0.8 (1.8)		-	0.8 (1.1)
NCI	CF	6.4 (0.7)	-	6.2 (1.3)	1.8 (0.9)	-	2.0 (2.1)
	Nbr		-	1.0 (1.6)		-	4.6 (3.5)
Uncued recall							
CI	CF	5.9 (1.2)	5.4 (2.1)	3.9 (2.2)	1.1 (0.3)	1.1 (0.2)	1.1 (0.3)
	Nbr			0.9 (1.3)		1.4 (2.0)	5.2 (3.1)
NCI	CF	6.3 (0.9)	5.7 (2.2)	5.1 (2.0)	2.3 (2.2)	1.0 (0.0)	1.3 (0.5)
	Nbr			1.8 (2.4)		0.7 (1.8)	3.4 (2.6)

Note. CI = category-implied list; NCI = no-category-implied list. Standard deviations are in parentheses.

As it is seen in Figure 13, the cued-recall participants showed a partial dissociation between memory and monitoring performance as a function of whether targets constructed categories or not. When the cued-recall participants studied the targets implying a category (such as a 'part of body' or a type of 'colour'), they could enhance their recall performance compared to the cued-recall no-category-implied participants: The context in which the targets were studied facilitated retrieval performance when targets encoded with their meanings which construct particular, discrete categories with other targets. Even though target items were exactly the same between the lists (CI and NCI lists), the cued-recall participants were able to construct higher-order units when such categorical relationships were encoded at the time of study. Although a full dissociation was expected, monitoring of CI target recall was only comparable to when such relationship was not (or could not be) utilised whereas retrieval was significantly greater in CI target recall than NCI target recall. This was attributed to observation that the target recall in category-implied target recall was almost perfect (92.5%), which thereby left a very low possibility to observe deteriorated monitoring performance (e.g., false recalls). However, contrary to the expectation, the dissociation disappeared in the uncued-recall group clearly: The uncued-recall CI and NCI participants had comparable retrieval performance as well as monitoring (AUC) performance.

6.10. Discussion

The results showed that the cued-recall participants used the category relatedness (inter-target association) at retrieval when such association is encoded at study. In other words, the context in which the target items are studied turned out to be an effective retrieval strategy, which enhanced retrieval and thereby resulted in comparable monitoring performance between category-implied list (CI) and no-category-implied list (NCI). Converging to this observation, when the cued-recall participants encoded the items in a semantic context that did not lead the participants to construct categorical relatedness between the targets at the time of encoding (NCI list), these cued-recall participants did not seem to utilise that kind of a categorical relatedness at retrieval, which solely existed amongst the retrieved targets normatively (indexed with lower memory quantity in NCI list target recall compared to CI list target recall amongst cued recall group). Expectation of a full dissociation in cued-recall group was, however, hindered by quite high retrieval performance in CI target recall, which left not much space to yield falsely remembered responses (e.g., incorrectly recognized interrelated candidates) to lower monitoring performance substantially.

However, the results of the experiment were counter to the expectation regarding uncued-recall group. It was found that the uncued-recall participants did not differ in terms of retrieval performance between CI (58%) and NCI study lists (64%) as well as their monitoring performance were comparable. This surprising finding was attributed to the following possibility.

The uncued-recall participants who studied NCI list might have realised that the targets they wrote down are somehow related to each other at the time of testing just like the uncued-recall participants who studied CI list. The inference that the uncued-recall participants who studied CI list utilised the categorical relatedness was still kept since, even when participants were provided with cues at testing, they had significantly higher retrieval performance in CI list compared to the cued-recall group which studied NCI list. Such source was utilised even by the cued-recall participants at retrieval time, so that, when participants were not provided any cues at testing, the advantage of categorical relatedness encoded at study would naturally be taken at testing, since the most effective source available to retrieve targets was this relatedness set up at study. However, the possibility that the uncued-recall group which studied NCI list had somehow realised that categorical relatedness between target items and used it retrieve the targets was thought to be quite low. Although this possibility could be quite low, it might be the case that the uncued-recall group which studied NCI list might have somehow used category relatedness not implied by the cues at study. Since the meanings of the targets studied in NCI list were non-dominant meanings of these targets, when the participants are not primed with any cues at testing (uncued-recall), uncued-recall participants could have turned to the original, dominant meaning of the targets. For instance, the meaning of the target word ‘green’ was implied as ‘a type of colour’ when it was paired with the cue word ‘grass’ (in CI list), however, its meaning changed to ‘not fully qualified’ when it was paired with the cue word ‘amateur’ (in NCI list). According to such as Merriam Webster English Dictionary, the former meaning of ‘green’ (type of colour) is listed as the first meaning of the word, whereas the latter meaning of it (not fully qualified) is the ninth meaning listed out of 10 available meanings of the word, green. Based on this possibility, the uncued-recall participants might have utilised the categorical-relatedness to ease retrieval of other related targets at testing by turning back to the dominant meanings of the targets (body parts or type of colours) as they were not provided with any cues at test to invoke the very same meanings of the targets that were implied at study. As a result, the uncued-recall groups which either studied CI or NCI lists might have used the category-relatedness at recall, which thereby resulted in the disappearance of the dissociation.

More importantly, however, the above-mentioned findings particularly on cued-recall confirmed that the participants in Experiment 4 did indeed utilise the categorical inter-target association in retrieval 'when such relatedness was encoded at study'. It should be noted that the guidance in retrieval via a restricted search set (e.g., 'search for colours') was inferred with the dissociation emerging between memory and monitoring performance, since when guidance is adopted to remember the targets belonging to some particular categories, the searching process is restricted to particular sets. Whereas that restriction could enhance target retrieval as the candidate items have higher probability to be the target items, recognizing the correct item (target) from amongst the candidates is difficult.

The inference drawn from the results was, in short, that the participants (cued-recall and uncued-recall groups which studied category-implied list) strategically utilised the inter-target association on the basis of category relatedness between the targets (as much as the number of instances belonging to a particular category increased in the study lists) rather than the retrieval based on only an automatic process which activate the retrieval of some other semantically-related target(s) via the targets already retrieved. In other words, the reason why the cued-recall participants could take advantage of the interrelatedness between the target items, such as 'eye, head, chest', and had higher retrieval performance (90%) compared to the cued-recall participants who studied the very same targets but in different contexts (62%), was that it was the targets associated to each other and encoded accordingly at the time of study. To be more specific, why would the target words 'eye, head, chest' etc. be retrieved almost 50% better when they were encoded in the 'body parts' contextual information than in contexts invoking the different meanings of them? This finding was attributed to the fact that participants encoded the meanings of these targets in the common contextual information (e.g., body parts or colours) so that they could utilise that information (as a cue) at retrieval. The following paragraphs will elaborate this issue further.

For instance, on the basis of free association norms of Nelson, McEvoy, and Schreiber, (1998), why would quite large number of people report that such as 'table' and 'chair' are strong/close associates (forward association: .76; backward association: .31) compared to the words such as 'king' and 'bicycle' (forward and backward associations: 0.0)? What would make the table-chair word pair so special to be decided as they are semantically close associates compared to 'king' and 'bicycle' pair? The probable answer to those questions seems that the semantically close associates have higher probability to belong to a 'closer available category' (share more commonality) than the weak semantically associate words in the semantic associative network (Bower, 1980). In order to explain this possibility, some theoretical approaches, which suggest specific models for the knowledge structure, as well as how the knowledge is represented in memory, will be reviewed in the following paragraphs.

The theories of knowledge structure can be classified into two broad categories: *associative network models* and *attribute models*. Associative network models mainly assume that the knowledge about the world is represented in a network in which the information is represented in the nodes of this network either in terms of hierarchical conceptual organizations (Teachable Language Comprehender, TLC, Model of Quillian, 1968, 1969) or the conceptual information is connected to each other with different lengths or distances in the network (Spread of Activation Model of Collins & Loftus, 1975) or each node in such network involves propositions (the smallest unit of knowledge) (Propositional Network Theories of i.e., Anderson, 1983). On the other hand, the attribute models suggest that the structure of knowledge in the cognitive system could be considered as that the concepts are the collection of attributes or features (Hunt & Ellis, 1999). Examples of these models are *Feature Set Theory* (Smith, Shoben & Rips, 1974), which proposes that the knowledge of concepts is a collection of features with 'defining features' (such as the features corresponding to define what a 'bird' is) and with 'characteristic features' (such as 'flying' property of a particular object in the 'birds category'; i.e., hummingbirds can fly, but ostriches can not) and *Prototype Theory* (e.g., Posner & Keele, 1970). Amongst these theoretical approaches to explain how knowledge is structured and how the information is represented in the cognitive system, there exists the spread of activation model of Collins & Loftus (1975). However, the propositions of prototype models to understand why some concepts, words or lexicons are close in terms of semantic association seem herein noteworthy.

According to the prototype theory (Posner & Keele, 1970), it is assumed that experiencing various instances of a category leads to an abstraction of the common attributes belonging to that category (i.e., airplanes). The concept pertaining to that class is represented with the abstraction of a characteristic feature, which is the best representation of a category: prototype of that particular category (such as a prototype of an airplane explaining how an airplane looks like). One of the proponents of the prototype theory was Eleanor Rosch, who proposed two classes of prototypes: one is germane to the 'perceptual information' and the other one is related to the 'semantic information' (Rosch, 1975, 1978; see also Rosch & Mervis, 1975). In relation to the current study, semantic prototypes are assumed to be connected to the propositions made for how the categories are formed in the cognitive structure (Rosch, 1978). According to Rosch (1978), there are two basic principles for the formation of categories. The first one is explained by the term *cognitive economy*, referring to the functional property of the category system that provides maximum information with least cognitive effort (for example, if categories have common features and are connected to the highest possible super-ordinate category; i.e., 'birds' and 'fishes' are represented under 'animals' category). The second principle is that the perceived world comes as structured information rather than as arbitrary and unpredictable attributes, and so the cognitive economy can be achieved either by the mapping of categories to the given attribute structures or by the definition of attributes rendering a given a set of appropriately structured categories. That is why, for instance, *correlational contingencies* are vital in the formation of category structures, such as 'wing' co-occurs with 'feather' but not with 'fur' so that it is parsimonious to categorise winged animals with feather (Rosch, 1978).

In short, the function of cognitive economy, as well as the importance of correlational (temporal and/or spatial) co-occurrence of the semantic information to form particular categories to which they belonged, seems to explain 'why semantically close associate words or concepts have a greater probability to construct a relatively closer category membership than semantically unrelated information'. In other words, thinking in terms of associative network models (such as anything might be associated to anything else in some way, either closely or remotely) as well as the semantic prototype application of Rosch mentioned above (1975, 1978), semantically strong associates are decided as strong associates since they have more common features to be combined to a 'closer available category' in the cognitive system. That may account for why people have a higher probability to report that 'table' and 'chair' are strong associates of each other rather than 'king' and 'bicycle'. That is, in the experiential repertoire of a person, chair and table have more temporal as well as spatial co-occurrence than king and bicycle have, so that the common attributes of chair and table seem to be connected with a closer higher-order unit, such as 'furniture', to which the decision as to whether the given concepts are linked to each other seems to be consulted.

Hence, the cued-recall participants in the present experiment seemed to construct higher-order units (categories) at encoding as long as the semantic context let them to do so. However, generating probable candidates on the basis of a restricted search set formed by this category relatedness would also involve a simple spread of activation. In other words, the strategic processing in retrieval does not simply rule out some automatic processes. Nevertheless, the activation process seems to be strategically guided according to the imposition of the list structure during study that turned out to be an established source or cue for retrieval. In this sense, the results of the experiment 6 indicated that the participants do not report targets 'only' on the basis of an automatic process rather they use a strategic searching process involving an automatic process of activation.

CHAPTER 7
General Discussion

The six experimental studies reported in this thesis primarily aimed to investigate memory performance as well as metacognitive processes in cued and uncued recall of paired-associates on the basis of the assumptions of generate-recognize model (e.g., Kintsh, 1968; Bahrick, 1969, 1970; Anderson & Bower, 1972). Specifically, however, I tried to answer two main research questions as follows.

First, is the assumption of generate-recognize theory that people generate more information than they report, and they withhold some of these responses due to stringent report criterion correct? In other words, can forcing participants to report the same number of items they study –with guessing when necessary- (e.g., forced-report) result in more correct responses compared to when participants are free to report as many items as they wish? More importantly, if that is so, under which conditions can that happen?

The second research question was based on the following assumption of the theory on recall process. When direct access fails, people are assumed to 'generate' some candidate items and then they make an attempt to 'recognize' the correct response (e.g., Bahrick, 1970). Based on this basic assumption of the theory and on which some strategies to measure metacognitive processes (e.g., monitoring, control, and report bias) via, such as strategic regulation of memory accuracy framework of Koriat and Goldsmith (1996c; Goldsmith & Koriat, 2008) and type-2 signal detection theory (e.g., Higham, 2002, 2011; also see Galvin, Podd, Drga, & Whitmore, 2003 for a review), the following was expected. When participants are put in a condition in which their memory search is restricted to particular sets in memory storage (e.g., categories), then the generated items are predicted to be inter-related. As a result of this, the recognition of the correct item (or response) was expected to be difficult. If this expectation, which is set up on the basis of the generate-recognize theory, is valid, we might expect a dissociation between memory and metacognitive monitoring performance. That is, providing that the participants are restricted in particular sets in memory search (e.g., by increasing the interrelatedness between to-be-remembered items), the generation process might be expected to be facilitated, and this facilitation would be observed with higher retrieval performance (e.g., forced-report memory quantity) since the to-be-remembered items would most probably be within these memory sets (e.g., categories).

However, as the generated responses would be related to each other, participants can be expected to have difficulty in recognising the correct item amongst these generated candidates, and so they report some incorrect responses as well. As a result of that, their monitoring performance, which refers to the ability to discriminate between correct and incorrect responses (e.g., Higham, 2002), was predicted to deteriorate, even though retrieval performance can be enhanced after studying inter-related target items. Based on this idea, the following question was asked: Does the dissociation between memory and metacognitive monitoring emerge as a result of studying inter-related items? Although it has been a well-known fact that inter-relatedness between target items (e.g., categorical relationship) increases memory performance (e.g., Tulving & Pearlstone, 1966; Cofer, Bruce, & Reicher, 1996; Runquist, 1970; Thomson, Hamlin, & Roenker, 1972), does it also have the same merit on metamemory or, on the contrary and as we expected, does it have a detrimental effect on such as metacognitive monitoring?

The following sections will review and discuss the results gathered in this research by answering the above-mentioned research questions in turn. After the general discussions of the results, the contributions of the current research and suggestions for future studies will be mentioned.

7.1. Does varying report criterion matter in observing what is actually remembered from memory (Experiments 1 & 2)?

The question as to whether variations in report criterion options have differential effects on measuring what is actually retrieved has been investigated in the contexts of such as hypermnesia (Roediger & Payne, 1983, 1985; Erdelyi, Finks, & Feigin-Pfau, 1989, Experiment 1 & 2; Roediger, Srivinas, & Waddil, 1989) and hypnosis (Dywan & Bowers, 1983; see also Klatzky & Erdelyi, 1985 for a theoretical discussion from signal detection theory perspective). The common finding of these investigations was that the conditions in which participants have no control over report option (e.g., forced-report) did not have a facilitative effect to yield higher number of correct responses compared to the those report conditions having stringent levels of report criterion (i.e., free-report, uninhibited-report criterions). In addition to this shared finding gathered when single-item study lists were used, Lockhart (1969), for instance, also observed that the reduced-report criterion did not have a facilitative effect on gathering higher number of correct responses in paired-associate learning as well.

As mentioned earlier in detail (see section 1.4. in Chapter 1), the generate-recognize theory assumes that there exists a two-staged process in recall in which people 'generate' some candidates and then they make an attempt to 'recognize' the target item from amongst the generated candidates. Notably, the theory was also taken into the stage of investigating such as the decisions as to whether a response would be reported or withheld (e.g., type-2 signal detection theory). As a result, the model was incorporated with a pivotal assumption that there exists a subjectively-set 'report-criterion', that is sensitive to the situational demands and/or pay-offs and is adopted by the participants so as to decide whether the response would be given or not (e.g., see Koriat & Goldsmith, 1996c; Higham, 2002, 2011). Hence, people might be expected to withhold some of the correct responses -if not all- that they generated and the proportion of those correct responses withheld or reported would vary as a function of the report option they are imposed by such as experimental contexts (i.e., free-report vs. forced-report) or the context in which the information is remembered and then reported (e.g., telling what happened last night's party to a friend vs. being an eyewitness in a court case). Besides, providing that report criterion does not vary such as between two particular conditions, then the differences between the percentages of reported and withheld items or responses can be attributed to differences in the distribution shifts amongst these two conditions. In some cases, however, the observations of memory and metamemory performance can also be attributed to the combination of distribution and criterion shifts.

Why is this question important? It is important mainly because a substantial number of studies have shown that report criterion does not much matter to observe what is truly retrieved, and forced-report as opposed to free-report conditions enhance the retrieval performance only with a small proportion of correct responses despite increasing the false alarms substantially (e.g., Roediger, Srivinas, & Waddil, 1989); see section 1.6.1. for detailed review on the subject. Hence, these results were taken as counter-evidence to the basic assumption of generate-recognize approach (e.g., Kintsch, 1968; Bahrick, 1970; Anderson & Bower, 1972). The failure of finding such enhancement in report criterions that are not controlled (e.g., forced-report) was highlighted with the following argument of Roediger and Payne (1985).

> "…the large variations in recall criterion produced by manipulating instructions at test (as measured by intrusions) did not affect the overall level of correct recall of the magnitude of improvement across tests. Apparently, the

assumptions of generate-recognize theories that people generate much more information in free recall than they produce (due to a stringent criterion for recognition of the generated material) is false." (p.1).

It should be noted herein that the present research had a basic motive so as not to simply abandon the generate-recognize perspective, rather it aimed to reveal the importance of using report-option variations not only in memory but also in metamemory research, just like some researchers (e.g., Koriat & Goldsmith, 1996c; Goldsmith & Koriat, 2008; Higham, 2002, 2011). In this sense, some recent efforts such as the strategic regulation of memory accuracy framework of Koriat and Goldsmith (1996c) as well as the measurement strategies of type-2 signal detection theory (e.g., Higham, 2002; Higham & Tam, 2005, 2006) which fundamentally base on the assumptions of generate-recognize models seem to be quite effective and functional tools which are currently available to investigate memory along with metacognitive processes in retrieval (although see Higham 2011, for a comparison of the available strategies to investigate metacognitive processes). Therefore, the current research targeted to investigate whether report-criterion variations change the pattern of observed retrieval performance, and if that might do so, under which conditions it happens.

Experiment 1, therefore, tested whether reduced report criterion results in higher number (or percentage) of correct responses compared to a condition in which participants have a relatively more lenient report-criterion. The participants studied randomly presented mixed list of 30 paired-words (target-cue), half of which had high semantic association (SA; e.g., *West - East*) and remaining half of them had low (or no) semantic association in the pairs (e.g., *West - Phone*). At the time of testing, the experimental-group participants were asked to remember and report 'only the target word' and the control-group participants were asked to remember and report 'all of the words they can remember'. These two groups were both uncued at testing and were asked to report their responses under free-report option. However, the control-group participants had a relatively relaxed (lenient) report criterion compared to experimental group due to the task demand at testing. The results of Experiment 1 indicated that reduced-report criterion (control group) yielded 'higher' number of correct responses compared to the group of participants who adopted a relatively more stringent report criterion (experimental group). This result was important to observe the fact that participants indeed withhold some of the correct responses that they had due to a relatively more conservative report criterion. Another implication of this central finding is that the performance measured when forced-report option is adopted approximates what is truly retrieved from the memory storage better compared to the conditions in which report-criterion is controlled (e.g., free-report; see e.g., Higham, 2002 for the argument on the importance of considering the control over report-criterion in memory research).

Then, Experiment 2 intended to find further evidence that forced-report option can result in higher percentage of correct responses compared to free-report option gathered in Experiment 1. However, it also targeted to investigate the effects of inter-target association (ITA) on memory and metamemory performance in cued and uncued recall of targets from the list of paired words (e.g., cue-target). The results of Experiment 2 revealed that forcing participants to report the same number of studied items (e.g., 24) resulted in significantly higher percentage of correct responses than letting participants report as many responses as they wished to report (free report), particularly when the cues studied are not overtly reinstated at testing (uncued-recall). This result was valid both in the uncued-recall of high-ITA targets and of low-ITA targets. However, finding higher memory quantity in forced-report uncued-recall group for high-ITA target recall compared to free-report uncued-recall group was inevitably expected, since participants could be able to utilise inter-target association as a source to remember the target items (e.g., via categorical relatedness) when they were forced to report. But, why did the forced-report participants have higher memory quantity in uncued-recall of low-ITA targets as well? This result seemed enigmatic at first, since there was no inter-target association amongst low-ITA targets to help retrieval of these unrelated targets when the uncued-recall participants were forced to report. The reason for this result was, however, attributed to the possibility of adopting highly stringent report criterion by the free-report uncued-recall participants in the testing time of low-ITA target recall. The combination of the following factors might have led to this possibility: The cues and targets were not related to each other neither categorically nor semantically, targets were not related to each other in terms of categorical association as in high-ITA target recall, and the participants were not provided with any covert cues at the time of testing. As a result, possibly a lowered feeling of learning (FOL) or feeling of knowing (FOK) (see e.g., Nelson, Gerler, & Narens, 1984; Koriat & Levy-Sadot, 2001) might have been reflected on reporting quite lower number of responses amongst the free-report uncued-recall participants (32%; e.g., see Table 6). In short, whereas the higher memory quantity observed in high-ITA target recall between free-report and forced-report uncued-recall groups was attributed to an available source (e.g., ITA) to ease retrieval of inter-related targets when participants were 'forced to report', higher memory quantity gathered for low-ITA target recall in forced-report uncued-recall group than free-report uncued-recall group was attributed to the relatively more stringent report criterion employed by the 'free-report uncued-recall' group when targets were unrelated. As a result, the free-report uncued-recall group strategically reported less in the recall of unrelated-targets for the sake of being more accurate in their responses.

The counter-evidence gained in Experiments 1 and 2 against some earlier studies of research that investigated the effects of report-option variations on measuring what is truly stored in memory was mainly attributed to the different procedures employed in the experiments. As mentioned earlier, the procedures in manipulating the report-criterion options can be classified into two main categories: the procedures that test report-option variations 'between participants' and those testing report-option variations 'within participants'. The lack of evidence observed in the studies which use between participants methodology to measure differential effects of report option on recall performance (e.g., Dywan & Bowers, 1983; Roediger & Payne, 1985) can be attributed to such as 'processing bias' (Erdelyi, 1985). In other words, participants might be dividing their total efforts across the items to be responded (e.g., cues) or put more effort onto the earlier responses (Erdelyi, Finks, & Feigin-Pfau, 1989). The second class of methodology, using within participants procedure, involves such as 'free report-draw a line-continue' methodology, (e.g., Erdelyi, Finks, & Feigin-Pfau, 1989) and 'report or pass' methodology (e.g., Higham, 2002; Higham & Tam, 2005, 2006); see section 1.6.1. for the details on the implementation of these procedures. In the 'free report-draw a line-continue' method, Cofer (e.g., 1967), for instance, suggested that participants might be expected to have already reported or dumped all the information they had in free report phase. As a result, providing that this possibility exists, expecting further correct responses to be reported under a subsequent forced-report condition seems unreasonable. Besides applying the possibility that participants may already deplete whatever correct responses they might have to within participants methodology (particularly 'free report-draw a line- continue' method), this possibility also exists for the between participants procedures, such as the ones comparing free-report condition to forced-report conditions.

Therefore, the results of Experiment 1 and 2 are hereof noteworthy. It was shown that higher percentage of correct responses can be gathered under less stringent or controlled report-option cases (e.g., control-group in Experiment 1 and forced-report groups in Experiment 2, respectively) compared to the conditions where more stringent report criteria are adopted (experimental-group in Experiment 1, and free-report groups in Experiment 2). This difference, then, can be achieved either by letting participants take advantage of some available sources (e.g., ITA) under forced-report option or by having participants adopt a quite stringent report criterion under free-report option which substantially hinders reporting all of the correct responses they had.

To sum up, people do indeed withhold some correct responses they have due to stringent report criterion (e.g., for the sake of being more accurate in their responses), and the differences on retrieval performance gathered between various levels of report options seem to stem from the over reliance of the utilization of contextual information (e.g., inter-target associations) under forced report and/or having the groups in which report-option is controlled (e.g., free report) to employ enough stringent report criterion.

7.2. Inter-relatedness between to-be-remembered items facilitates retrieval, but what about its effect on metamemory (Experiments 1-6)?

Inter-relatedness amongst the items studied (e.g., categorical relations) has been found to have a facilitative effect on remembering these items (e.g., Cofer, Bruce, & Reicher, 1996; Runquist, 1970; Thomson, Hamlin, & Roenker, 1972). However, its effects on metamemory have remained unclear. Hence, Experiments 1-6 tried to answer the following question: Does inter-relatedness between targets have the same merit such as on metacognitive monitoring as it has on memory performance?

Experiment 1 showed that whilst participants recalled higher percentage (MQ) of high-SA targets than low-SA targets, they had comparable percentages of memory accuracy in high-SA target recall to low-SA target recall both in experimental and control groups.[28] That is, the percentage of 'incorrect responses' reported voluntarily (since both experimental and control groups were tested under free-report) in high-SA target recall was significantly higher than in low-SA target recall. As a result, participants seemed to be monitoring their responses in high-SA target recall less

[28] The study list was mixed in Experiment 1. Therefore, the nature of responses given that were neither cues nor targets could not be separated as to whether these responses were given as semantic associates of high or low-SA targets. Although available normative data would let us understand the activation source, considering anything might be activated by anything else depending on the experiential repertoire of participants, these responses were treated exactly the same as they were equally activated by high and low-SA pairs. Hence, a prospective study that intends to use the design of Experiment 1 is recommended to use separate lists, involving only one type of study list material (e.g., high-SA or low-SA pairs) in each study list.

efficiently as in low-SA target recall. Besides, as a type-2 signal detection measurement, their AUC scores did not differ between high and low-SA target recall as well. Manipulating inter-target association (high vs. low) via using exemplars of categories to be target items, Experiment 2 also substantiated this pattern. Particularly uncued recall participants had higher memory quantities where ITA was taken as a retrieval source, although their monitoring performance (e.g., AUC scores) did not differ (partial dissociation). Experiment 3 which used within participants report-option (report-or-pass methodology) again revealed that particularly uncued-recall participants rather than the cued-recall participants utilise the existing ITA when the pairs are encoded with target-focused way, but not when the pairs were studied by focusing on the pairs as whole. In other words, study context was a leading factor that had participants to utilise inter-relatedness between targets at retrieval, and so differential effects of ITA on memory and metamemory (e.g., AUC) could be observed (e.g., partial dissociation). However, besides the fact that the manipulation in ITA level did not have any effect on memory so that on metacognitive monitoring in cued-recall, the retrieval (as well as monitoring) was not affected by the variations in study context (e.g., target-focused vs. pair-focused study types) and in retrieval context (e.g., liberated vs. constrained in responding) amongst cued-recall group as much as uncued-recall (e.g., target-focused study yielded partial dissociation in uncued-recall group). This pattern implied that cues are effective sources in recall and other sources, such as ITA, do not seem to be an additive factor in cued-recall providing that the cue items are encoded well enough to elicit the target paired with it at retrieval.

Because the number of categories to which target items belonged in Experiment 3 was considered as a restricting factor to lead the expected full dissociation between memory and metamemory as a function of using ITA, the number of categories in Experiment 4 was reduced from four down to two categories and memory and metamemory performance in two-category study list was compared to the performance in six-category and in twenty-four category study lists. Experiment 4 showed a full dissociation between memory and metamemory performance which emerged as a result of utilising inter-target association both in cued-recall and uncued-recall groups. This finding was gathered on the basis of the basic premise of generate-recognize models: When generation process is facilitated, the generated items (responses) turn out to be inter-related, so that recognizing the correct responses from amongst the related candidates becomes difficult (see e.g., Higham & Tam, 2005 for the argument). This dissociation is critical to demonstrate the following. Memory enhancement methods such as increasing the categorical interrelatedness between target items seem to come with a cost on metacognitive monitoring performance. Converging to the findings of Guerin and Miller (2008), Experiment 4 substantiated that the inter-relatedness amongst to-be-remembered items has a dissociative effect on recall and recognition, due to the mixed blessing nature of inter-target relatedness. The facilitation in generation process was inferred by forced-report memory quantity performance, whereas deterioration in the recognition process was observed with response-contingent monitoring. Different to the study of Guerin and Miller, the dissociation was shown in a 'single' retrieval process (recall) rather than comparing two distinct retrieval processes (recall and recognition), and in a paired-associate learning experiment, which gave us a higher chance to observe under which conditions the reinstated cues play a central role in retrieval against, such as inter-target association.

Following Experiment 4, Experiment 5 manipulated the encoding processes (interactive imagery vs. rote repetition) so as to test whether the dissociation can be attenuated when the pairs are 'individuated' enough. This testing was also important to show that the dissociation found in Experiment 4 does indeed emerge when ITA is encouraged to be utilised in guiding memory searching process, which thereby enhances recall performance but deteriorates monitoring performance of the responses at the same time. The dissociation was diminished amongst cued-recall participants via using an encoding strategy that leads to a more relational mnemonic, which was interactive imagery. However, dissociation was still observed when cues are not reinstated at retrieval (uncued recall). This pattern was important to show us the fact that when participants are not presented with any cues (together with the condition in which the cues are not helpful to ease retrieval of their paired targets via strong semantic association), uncued-recall participants seem to take advantage of the most efficient and available source to retrieve the target items such as ITA. In other words, the knowledge gained about the study list structure (e.g., ITA existence) comes about a central factor to play a role in retrieving to-be-remembered items in uncued-recall, if remembered cue words are not taken into consideration to retrieve the target items they were paired with.

Experiment 6, which used homograph words, showed that cued-recall group yielded the expected dissociation between memory and monitoring performance due to the utilisation of inter-target association at retrieval when such relatedness was implied by the cues at study. Counter to the expectation, the dissociation was not observed in uncued-recall group. This result was attributed to the possibility that the uncued-recall participants might have used the categorical-relatedness to ease retrieval of other related targets at testing by turning back to the dominant meanings of the targets. Hence, the uncued-recall groups, which studied either category-implied list or no-category-implied list, might have used the category-relatedness at recall no matter what the cue words implied at study, and so the dissociation disappeared in uncued-recall group.

Experiment 6, which used homograph words, showed that cued-recall group yielded a partial dissociation between memory and monitoring performance due to the utilisation of inter-target association at retrieval when such relatedness was implied by the cues at study. Counter to the expectation, the dissociation was not observed in uncued-recall group. This pattern was attributed to the possibility that the uncued-recall participants might have used the categorical-relatedness to ease retrieval of other related targets at testing by turning back to the dominant meanings of the targets. Hence, the uncued-recall groups, which studied either category-implied list or no-category-implied list, might have used the category-relatedness at recall no matter what the cue words implied at study.

The results gained in Experiment 6 was also important to indicate that when generate-recognize route is in process, it is not ‘solely’ a simple, mechanistic, and automatic activation process between ‘normatively’ related targets, but (although incorporating some automatic processes as well) it seems that there exists a process emerging from the ‘sensitivity to the study list structure and guides the memory search accordingly and strategically’. In this sense, a radical change that was made by Jacoby and Hollingshead (1990) in the early versions of generate-recognize models (e.g. Atkinson & Shiffrin, 1968; Bahrick, 1970; Anderson & Bower, 1972) seems valid. Put it differently, the source of generating candidates is a distributed rather than a stable associative network (episodic memory) rather than an abstract and stable network in (semantic memory) (e.g., see Bower, 1980; see also Higham & Tam, 2005 for a review of early and modern versions of generate-recognize models). Following this assumption, particularly the results of Experiment 6 showed that the knowledge gained about the study list structure as episodic information becomes a critical factor at generation process that may surely be triggering some automatic processes as well (e.g., spreading of activations; cf. Collins & Loftus, 1975). Besides Experiment 6, the results of Experiment 5 also converged to the observation that knowledge about list structure, which is guiding the memory search strategically, turns out to be a leading factor to retrieve memories. In this vein, dissociation was observed amongst uncued-recall participants regardless of the encoding strategy employed.

In short, the current research showed that increasing inter-relatedness between target materials facilitates memory, but it does not have the same merit on metacognitive monitoring, and the monitoring is deteriorated to the extent that inter-target relatedness guides memory retrieval.

The following paragraphs, which based on the assumptions and measurement strategies of type-2 signal detection theory will try to clarify why the dissociation was expected and how it was confirmed in this research.

Let me consider the empirical finding such as in Experiment 4, which clearly showed a full dissociation between memory performance and metacognitive monitoring due to high inter-relatedness between target items. The expectations in distribution and/or criterion shifts in cued-recall and uncued-recall groups as a function of number of categories in the study lists are displayed in Figure 15. In order to depict the patterns of distribution and/or report-criterion shifts clearer in cued-recall recall, the response-contingent hit rates (HRs) and false alarm rates (FARs) of the cued-recall groups were those based on strict scoring, instead of liberal one. The cued-recall group was found to have equal HRs between two-category and 24-category target recalls (M = .97, SD = .04; M = .98, SD = 0.7, respectively), $t(28) = -.119$, $p > .05$; However, FAR was significantly higher in two-category target recall than 24-category target recall (M = .38, SD = .29; M = .22, SD = .27, respectively), $t(29) = 3.079$, $p = .005$. In other words, although HRs were comparable, FARs differed between the recalls of related and unrelated items. Considering that inter-related items (e.g., two-category study list) are highly probably to activate other related items or responses (false signals) compared to unrelated or distinctive items (e.g., 24-category study list), the distributions of correct and incorrect responses generated in inter-related targets recall are expected to be closer to each other compared to those of unrelated-targets recall. Therefore, the results showing comparable HRs but higher FAR in related-targets recalls imply that there existed a shift in report criterion in cued-recall group. In other words, the cued-recall participants in Experiment 4 seemed to have relatively more liberal report criterion when targets were inter-related compared to when targets have no such inter-relatedness at all.

The situation in uncued-recall group was, however, different than the cued-recall group in Experiment 4. The comparisons of HR and FAR showed the following. The HR was significantly lower in two-category target recall (M = .87, SD = .21) than that in 24-category target recall (M = .93, SD = .17), $t(29) = -2.367$, $p = .025$. Also, FAR in two-category target recall was higher (M = .38, SD = .34) than that in 24-category target recall (*M* = .27, *SD* = .26), *t*(29) = 2.239, *p* = .033. Firstly, this pattern showed that monitoring performance was negatively affected more in uncued-recall compared to cued-recall groups observed in the current research (e.g., Experiments 3 & 4), mostly because the categorical-relatedness was utilised to the greatest extend in free recall (uncued-recall groups) than cued-recall. There seemed to be a distribution shift rather than a shift in report criterion in uncued-recall by considering the same expectation that the distributions of correct and incorrect responses are further apart compared to the distributions of studying less distinctive items (e.g., two-category study list) when items (in this case, responses) are distinctive along with considering the patterns of HRs and FARs. Hence, although there might have existed a slight criterion shift in uncued-recall, this was greater in cued-recall compared to uncued-recall. This situation that is displayed in Figure 15 converges with the observations in the current research as well as some earlier findings (e.g., see Bousfield, 1953; Cohen, 1966; Jenkins, Milk & Russel, 1958; Tulving, 1962): The category-relatedness is taken more of a retrieval source in generating responses when participants are not provided any cues at test (and when the cues are not strong associates of targets) than when participants are provided with cue items at test. As a result, when inter-target association is taken as a quite solid base to retrieve target items particularly when the participants are not given any overt cues at test, the distributions of correct and incorrect items are less and less separated. In other words, other related responses are generated to the highest extend as much as existing inter-relatedness between target items is employed to guide retrieval process.

7.3. Contributions of the present research and suggestions for future investigations

The central contribution of the current research is to support the expectations built on the primary assumptions of generate-recognize models. It was substantiated that when people have control over report criterion, which allows making a decision to report or withhold the responses that they have, people have a tendency to withhold some responses (such as the correct ones) as a function of the level of report-option adopted. Further, it was shown that the inter-relatedness existing amongst the to-be-remembered items turns out to be a trade-off: Although it enhances memory, it might come with a cost on metamemory processes such as monitoring (dissociation). Related to the latter finding, observing such dissociation was achieved on the basis of type-2 (response-contingent) signal detection theory. Therefore, the merits of type-2 signal detection theory to measure memory together with metacognitive processes was also

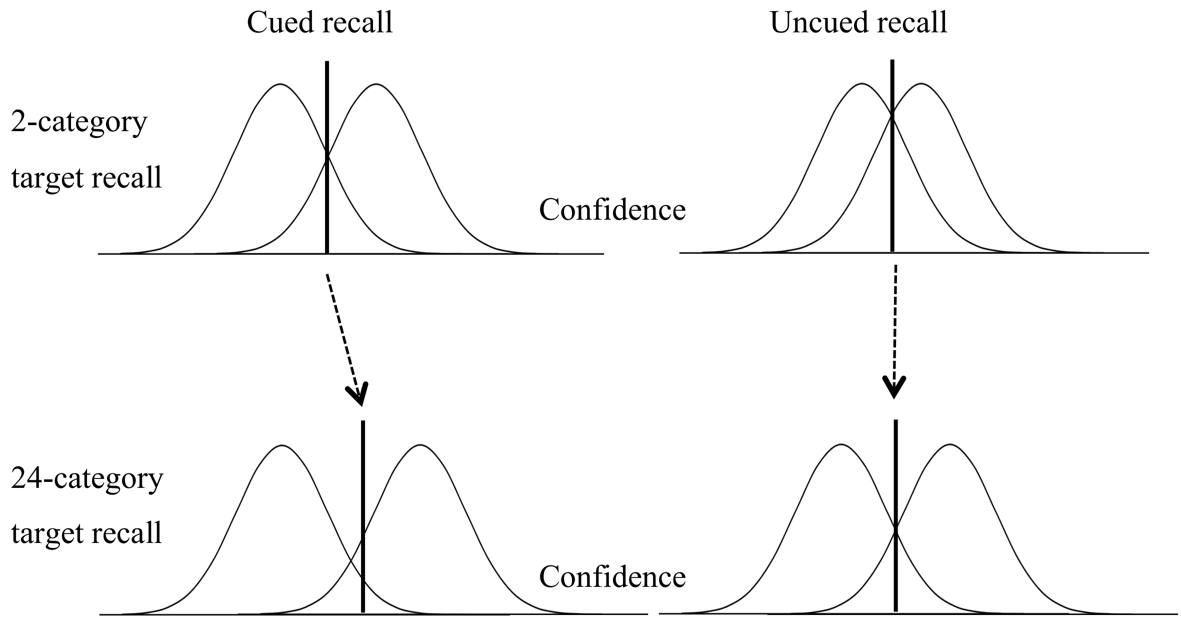


Figure 15. Hypothetical criterion and distribution shifts in two-category and 24-category target recall amongst the cued-recall and uncued-recall groups of Experiment 4. Note that hit and false alarm rates are response-contingent (based on type-2 signal detection modelling). Whereas the left-side distribution in each of the four paired distributions are of the distribution of the incorrect responses, the right-side ones are of the correct responses generated. Vertical lines display report criterions.

underlined via supporting the expectations that were built on the basis of the generate-recognize approach (see e.g., Higham, 2011 for a theoretical discussion and comparisons on the available methods to investigate metamemory). Since measuring metacognitive processes seems to be a more difficult job compared to, for instance, observing memory performance per se, type-2 signal detection theory seems to offer a functional tool to empirically observe what is genuinely going on in cognitive as well as in metacognitive processes.

To be more specific, however, the results gained in the current thesis seem to take some implications from various studies of research. These investigation fields are the ones on such as how metacognitive judgements (i.e., judgements of learning) are made, false memory, source monitoring, retrieval-induced forgetting as well as the neurological studies that investigate both memory and metamemory performance.

With regards to the judgements of learning (JOLs), for instance, the *cue-utilization approach* that is basically applied to how judgements of learning (JOLs) are made (Koriat, 1997) seems hereof so much related to the results of the current thesis. According to this approach, JOLs, like other metacognitive judgements such as feelings of knowing (FOKs) and subjective confidence are assumed to be 'inferential' in nature. Hence, the approach has been contrasted to the *direct-access hypothesis* (cf. King, Zechmeister, & Shaughnessy, 1980), assuming that participants can 'directly monitor trace strength' and make their JOLs on the basis of the assessed memory strength of the sought-after items. More important, three general modes of cues exist in the approach as Koriat proposed: 'intrinsic', 'extrinsic', and 'mnemonic' cues. Whereas *intrinsic cues* can involve the characteristic of study items (i.e., perceived or pre-experimental difficulty or ease of items, degree of associated relatedness between cue and targets, imagery values of the items, and so on), *extrinsic cues* are proposed to be the learning conditions (i.e., presentation time, massed or distributed learning, etc.) or the encoding operations employed by the rememberer (i.e., levels of processing). The third class comprising of *mnemonics*, that signal the extent to which an item has been learned and will be recalled in the future for the participants (i.e., internal, subjective, and phenomenal experiences such as retrospective confidence, the ease of information retrieval, the memory for the outcome of previous recall attempts). Based on the cue-utilization approach of Koriat (1997), the participants in the present thesis seemed to monitor their responses poorly to the extent that they rely on utilizing categorical relatedness, in this case 'intrinsic cues' (e.g., inter-target association), more than the mnemonic cues. Being non-mnemonic cues, inter-target association level and the knowledge which could be gained about the list structure tend to come along as a central source for the participants so as to both guide the retrieval process towards circumscribed category sets and to activate other related items (candidates) within the categories to which the searching process targeted. The interrelated candidates generated as a result of the meta-knowledge set up about the list structure, then, seemed to be falsely reported (rather than correctly withheld), since the non-mnemonic cues (rather than the mnemonic ones, e.g., intrinsic ones) were heavily relied on.

But, what could happen if the participants had been allowed to remember and report their responses at a later time? Would they still fall in to trap of recognition difficulty due to the increased number of highly related candidates generated? The answer to this question seems to lie on the results of the studies dealing with *delayed judgements of learning* (delayed JOLs; e.g., Dunlosky & Nelson, 1994). Delayed JOLs studies overall showed that when participants make their learning judgements later rather than immediately after studying lists, they are more accurate in their judgements (e.g., see Rhodes & Tauber, 2011 for a meta-analysis of delayed judgements of learning studies). The critical finding showing that 'delayed JOLs are more diagnostic of future recall performance' seem to imply that it is mainly because the participants rely more on mnemonic cues rather than non-mnemonic ones when they are allowed to form them via having enough time. Specifically, the delay has been proposed to let participants refrain from depending on highly probable false signals leaking from short-term memory (STM) and allows them to consider the information form long term memory (see e.g., Nelson & Dunlosky, 1991, Dunlosky & Nelson, 1992, 1994).

The future investigations, therefore, can be quite instructive by answering to what extend the possible dissociation between memory and monitoring performance coming along due to heavier reliance on mnemonic cues (e.g., categorical/semantic relatedness between target items) could be attenuated when such as retrospective confidence judgements are made at a later time. Besides, the future investigations could incorporate judgements of learning into the current experimental designs unlike the experiments reported in this thesis. Incorporating JOLs into the current experiments could shed more light onto the possibility that whereas the fluency heuristic seems to let participants generate from the correct category sets, it seems to overshadow the distinctive features of generated responses. Therefore, the prospective studies could also converge to the present implications by showing that the dissociation largely comes about due to heavier reliance of inter-target association in the study lists and this reliance could diminish by time and give its place to mnemonic cues, which turn out to be more diagnostic and healthier in making such as metacognitive judgements (e.g., retrospective confidence judgements).

The cue utilization approach of Koriat (1997) prosed to be applied not only to judgements of learning but also to any kind of metacognitive judgements, connotes other investigations as well. These investigations amongst them such as those dealing with false memory, source monitoring and retrieval-induced forgetting, however, seem hereof more noteworthy. The following paragraphs will firstly mention false memory research together with an examination of the closest available example as well as a well-known one in the literature - the Deese/Roediger-McDermott effect (Deese, 1959; Roediger & McDermott, 1995) with respect to the results gained in the current thesis and then, it will reach to the possible implications on source monitoring and retrieval induced forgetting studies.

The dissociation gathered in the present research seems to be critical to contribute to the understanding in false memory research as well as take some explanations from these studies. The investigations on the Deese/Roediger-McDermott (DRM) effect, that was named after Deese (1959) and Roediger and McDermott (1995) and created as a memory illusion in laboratory setting (i.e., false recall or recognition of a 'critical lure', *needle*, after studying the words, *thread, pin, eye, sewing, sharp,* and *point*), are herein necessary to be mentioned. Although the experiments in the current research primarily investigated a type of false memory in recall via assuming existence of two successive processes, which are generation and recognition, some resemblance can be made between the present research's ideas and the explanations as to why DRM effect is observed in terms of the failures in source monitoring. The following paragraphs will elaborate on the subject in detail.

DRM effect has been mainly explained by *activation/monitoring hypothesis* (Roediger, Watson, McDermott, & Gallo, 2001), in which the critical lures are thought to be activated particularly at study time and people have a difficulty in 'monitoring the source of this activation' (i.e., from the study lists or they are self-generated). The activation process is suggested to be affected by top-down processes (i.e., associations, gist, inferences, categories, etc.) as well as by bottom-up processes (i.e., feature overlap, feelings of familiarity, etc.). The monitoring, on the other hand, describes any decisional and editing processes that aid determining the origin (source) of the activated information. Hence, this most recent understanding, which has been suggested to be applicable not only to the DRM effects but also to various false memories by Gallo (2010), emphasizes the notion of spread of activation amongst pre-existing conceptual representations in mental lexicon (e.g., Collins, & Loftus, 1975). That is why DRM effect has been given as a solid example for associative memory illusions by many scholars (i.e., Park, Shobe, & Kihlstrom, 2005; Dewhurts, Bould, Knott, & Thorley, 2009; see also Roediger, & McDermott, 1995; Smith, Gerkens, Pierce, & Choi, 2002; Knott & Dewhurst, 2007; Dewhurst, Bould, Knott, & Thorley, 2009).

Therefore, the reason as to why inter-related items were expected to facilitate generation process, but deteriorate recognition of these items in the present thesis becomes clearer by considering some explanations and findings in false memory (e.g., DRM effect). First, expecting inter-relatedness between items facilitates the generation process becomes understandable in the light of the studies which clearly show categorical or semantic associations amongst to-be-remembered items enhance retrieval (e.g., Tulving & Pearlstone, 1966; Runquist, 1970; Thomson, Hamlin, & Roenker, 1972) as well as with the findings that activation process is boosted by inter-relatedness amongst target items (e.g., Roediger & McDermott, 1995; Roediger, Watson, McDermott, & Gallo, 2001; Park, Shobe, & Kihlstrom, 2005; Dewhurts, Bould, Knott, & Thorley, 2009; Gallo, 2010). The second expectation that predicted a difficulty in recognition process (so that a type of false memory could be observed) due to facilitated generation process seems more related to the explanations made on 'the relationships between activation and monitoring'. That is, the participants seemed to have poor monitoring with regards to the source of the retrieved items (e.g., from the study lists vs. self generated). As a result of this, they tended to report some false responses, which had less distinctive features amongst the generated candidates, along with substantial amount of correct responses (dissociation or a trade off). Herein, *retrieval-induced forgetting* (e.g., Ciranni & Shimamura, 1999) does not seem at odds with the implications gathered in this thesis as well. This phenomenon refers to the happening in which the retrieval of related items (e.g., concepts, words, geometrical shapes, etc.) compete with the sought-after item, and the intended item becomes available only after the residual activity amongst the incorrectly retrieved (or activated) items has decayed. When failures in source memory as well as retrieval-induced forgetting findings are considered together, the mixed blessing of high inter-target association observed as a dissociation between memory and metacognitive monitoring performance seemed to appear as participants generated related candidates after a strategically guided searching process so that the inter-related responses competed with each other and they resulted in forgetting the inter-related but correct responses. As a result, participants seemed to monitor the source of the responses 'poorly' and reported some falsely generated responses as well due to the less distinctive features existed amongst generated candidates. Although Experiment 5 gave a hint that the possible failures in source monitoring could be attenuated by observing the dissociation diminishes as a result of individuated pairs (particularly in cued recall), the prospective studies can be so much informative if they directly target at substantiating the existence of a possible failure in source monitoring that is conducive to the dissociation in the present experimental designs. One possible type of study can compare the recognition of categorically related words and their visual depictions (e.g., pictures) between participants and could expect a possible dissociation between memory and monitoring performance when study lists composed of words rather than pictures (assuming that the pictorial information could inevitably have greater number of distinctive features than, for instance, their verbal or conceptual counterparts).

As a practical application of theoretical explanations, source monitoring has also been investigated in eyewitness memory (e.g., Multhaup, De Leonardis, & Johnson, 1999). Although making suggestions on how court cases, cognitive interviews, and/or eyewitness testimonies should be run in order to gather 'all' but the 'accurate' information about the events witnessed should be approached cautiously, the findings gathered in the present thesis seem to imply some suggestions regarding the eyewitness testimonies where recognizing the offenders amongst the suspicious ones is the case. For instance, consider the following hypothetical example. A person who witnessed an event happened one day before states that he remembers there were three offenders involved in the event and all of them were male, tall, and around 30. When the suspects amongst whom the offenders are to be detected are all from the intersection set (just like the searching process guides memory to the restricted category sets due to the existence of high inter-relatedness amongst to-be-remembered items), the probability of detecting the real offenders seems to be greater compared to when the suspects are from various different categories so that they are almost impossible to be real offenders (e.g., all of the suspects compose of females, children, and elderly). However, recognizing the real suspects amongst the very similar ones could be endangered as the discriminatory features weaken. As a result of this, the chance of detecting one (or several) suspect would be high, although detecting some other suspects falsely would be the case as well. In short, confidently-remembered features which designate some commonalities (so that restricting the memory search sets) should be the ones that are highly probable to be confidently remembered ones (e.g., 30 years old tall males), but they should also be distinctive enough at the same time to yield ‘efficient memory’ as well as ‘accurate recognition’. Hereof, the following finding of Perfect and Schwartz (2002) has major implications regarding the above-mentioned example: Whereas personality types do not predict much how eyewitness accuracy, people who believe they have strong memories often express overconfidence leading to poor memory accuracy. That is, some factors or beliefs stating a sound functioning of memory (just like the very existence of inter-target associations in the study lists) could lead to reduced memory accuracy. In this sense, the present research suggests that some memory-facilitating techniques (e.g., categorical and/or semantic relationships amongst target items, and even method of loci –providing that it is ‘overly’ used) can endanger metacognitive monitoring performance due to weak distinctive features amongst the to-be-items to let memory be able to monitor the sources of information (e.g., correctly remembered or falsely generated).

Last but not least, the neurological studies investigating memory along with metacognitive functioning in various clinical populations are also much informative on the implications gathered in present research. For instance, Moulin, Perfect, and Jones (2000a; see also 2000b & 2000c) investigated metacognitive judgements made in

Alzheimer's disease (AD), which is encountered as the most prevalent type of dementia (Wimo & Prince, 2011). Moulin and his colleagues (2000a) compared JOLs in AD group and age-matched controls, and they found that these groups did not differ in terms of accuracy scores of their JOLs. However, AD patients' accuracy scores did not differ from zero, which implied that there exists some level of difficulty in predicting the likelihood of remembering the stored information amongst AD patients. Unlike the memory problems, which are quite apparent in such as Alzheimer's disease, existence of monitoring deficits in this kind of neurological disease has not been shown markedly (Moulin, 2002). In better relation to this as well as to the implications of the current research, the investigation of Moulin, Perfect, Conway, North, Jones, & James (2002) who compared retrieval-induced forgetting in Alzheimer's disease patients and age-matched controls is herein notable. Moulin et al., (2002) found quite interesting evidence that the inhibition of items in episodic memory was comparable between AD patients and their age-matched controls, which indicated that inhibition of competing items (e.g., unpractised but relevant items) is intact in AD. The general implication of this study indeed pointed out a possibility that the retrieval-induced forgetting observations do not seem to be resulted from deficits in inhibitory processes (Moulin et al., 2002). As they pointed out, intrusion errors are not necessarily a result of inhibition failure, which implies that such errors might be made on the basis of gist, or a failure in source monitoring (e.g., see Budson, Daffner, Desikan, & Schacter, 2000; Multhaup & Balota, 1997; see also *Fuzzy Trace Theory*, Reyna, & Brainerd, 1995, for the distinction between direct access to verbatim traces vs. reconstructive processing of gist traces). As an interpretation of the finding, suggesting a possibility that intrusion errors do not necessarily stem from inefficient functioning of inhibitory processes, perfectly well functioning inhibitory processes might be observed at best in the cases where the responses, items, or candidates 'mismatch' to the context (composed of such as beliefs, expectations, generated hypotheses, etc. as a package) rather than the ones where there exist modestly matching items (i.e., retrieval of unpractised but related items, RP^-). Given that the above-mentioned logic is valid, the studies which aim at finding possible reasons as to why intrusions errors are observed (e.g., due to deficits in inhibitory processes, failures in source monitoring, heavy dependence on verbatim traces, etc.) might consider the following. For instance, inhibitory functioning could be interpreted clearer when such a functioning is assumed to vary on a continuum (ranging between perfectly-well inhibitory process, e.g., everyday inhibition or repression of the memories which are out of the context and extremely poor inhibitory process, such as flash bulb memories), rather than it is interpreted categorically (e.g., inhibition process is efficient or not).

To sum up, it was shown as a central implication of the present thesis that the proposed interdependency between cognitive and metacognitive processes (e.g., see section 1.2. and Figure 1) does not seem to function well 'at all times'. For instance, some acquired metacognitive knowledge and/or inferences taken by episodic occurrences (i.e., 'the targets I am asked to report are inter-related in some way') seem to strategically guide the searching process at retrieval. However, the directed guidance in searching process falls into the trap of recognition difficulty most probably due to the failures in source monitoring (e.g., whether the generated responses are indeed from the study list or 'produced' only at testing; see e.g., Johnson, Hashtroudi, & Lindsay, 1993). Hence, the findings in the present research yielding a clear dissociation between memory and metamemory performance due to high inter-relatedness amongst targets show that some variables (e.g., categorical/semantic associations amongst to-be-remembered items) have differential effects on cognitive and metacognitive levels of memory (which refers to object-level and meta-level of Nelson & Naren's (1990) framework, respectively), and so it seems that efficient functioning of one level does not always necessitate a subsequent healthy functioning of the other level. This main implication of the current thesis, then, can be extended to other fields that mainly focus on the importance of monitoring performance rather than gathering better retrieval performance as a priority, or at least, to those investigations which give equal importance to both parts of the performance. Therefore, as Gallo (2010) suggested that the accumulated knowledge in false memories should be generalized outside laboratory, the importance of 'distinctiveness' in monitoring performance observed in the present research can be extended to the investigations of such as eye-witness testimonies, cognitive interviews as well as the investigations in educational and clinical settings in which 'the accuracy of the responses or the monitoring of them is more critical' than facilitating retrieval of sough-after information per se.

APPENDICES

Appendix A

Materials used in Experiment 1

Table 18

The Target and Cue Words, and the Forward and the Backward Semantic Association (SA)Between the Targets and Their High-Semantically Associated Cues

			List-1					
			High-SA pairs					
#	Targets	Cues	C.cn.	C.fr.	T.cn.	T.fr.	F.A.	B.A.
1	Bulb	Light	6.12	0	5.39	333	79	21
2	Calculus	Math	3.72	0	3.78	4	71	76
3	Movie	Film	5.85	29	5.91	96	19	54
4	Time	Clock	3.79	1599	6.94	20	37	65
5	Hammer	Nail	5.77	9	5.42	28	80	62
6	Test	Exam	5.08	119	5.14	29	25	78
7	Mother	Father	5.47	216	5.90	183	60	71
8	Cat	Dog	6.21	23	5.75	75	51	67
9	Garbage	Trash	5.68	7	5.76	2	46	53
10	King	Queen	-	-	6.38	41	77	73
11	Leaf	Tree	5.89	12	6.62	16	64	16
12	Gas	Fuel	5.34	98	5.47	17	38	66
13	Noun	Verb	3.93	1	3.17	4	64	69
14	Odour	Smell	5.83	14	4.40	34	70	16
15	Pepper	Salt	5.61	13	5.69	46	69	70
			Low-SA pairs					
16	Street	Mineral	5.84	244	0	12	-	-
17	Earth	Bargain	5.77	150	2.71	7	-	-
18	Pony	Magazine	6.51	10	-	-	-	-
19	Picture	Citizen	6.75	162	-	-	-	-
20	Grass	Shampoo	5.93	53	0.00	2	-	-
21	Square	Garlic	5.15	143	6.23	4	-	-
22	Beach	Blanket	6.08	61	6.21	30	-	-
23	Fork	Island	5.25	14	6.40	167	-	-
24	Man	Diamond	6.14	1207	6.21	8	-	-
25	Table	Yellow	6.00	198	5.18	55	-	-
26	Tale	Holiday	3.39	21	3.60	17	-	-
27	Sweet	Dancer	4.53	70	-	-	-	-
28	Rose	Jacket	5.86	86	6.31	33	-	-
29	Present	Muscle	3.37	377	6.00	42	-	-
30	West	Phone	4.03	235	6.02	54	-	-

Table 18 (*continued*).

	List-2							
	Low-SA pairs							
	Targets	Cues	C.cn.	C.fr.	T.cn.	T.fr.	F.A.	B.A.
1	Bulb	Piano	6.12	0	6.26	38	-	-
2	Calculus	Military	3.72	0	0.00	212	-	-
3	Movie	Plastic	5.85	29	0.00	31	-	-
4	Time	Dwarf	3.79	1599	-	-	-	-
5	Hammer	Pupil	5.77	9	5.66	20	-	-
6	Test	Drum	5.08	119	-	-	-	-
7	Mother	Crazy	5.47	216	-	-	-	-
8	Cat	Pudding	6.21	23	4.41	0	-	-
9	Garbage	Play	5.68	7	4.22	200	-	-
10	King	Traffic	-	-	5.83	68	-	-
11	Leaf	Lens	5.89	12	5.52	12	-	-
12	Gas	Library	5.34	98	5.30	62	-	-
13	Noun	Ghost	3.93	1	4.25	11	-	-
14	Odour	Leader	5.83	14	5.83	74	-	-
15	Pepper	Science	5.61	13	4.14	131	-	-
	High-SA pairs							
1	Street	Road	5.84	244	5.48	197	31	35
2	Earth	Planet	5.77	150	5.64	21	16	61
3	Pony	Horse	6.51	10	6.03	117	75	11
4	Picture	Frame	6.75	162	6.00	74	32	81
5	Grass	Green	5.93	53	5.46	116	36	25
6	Square	Circle	5.15	143	5.14	60	47	63
7	Beach	Sand	6.08	61	6.25	28	39	72
8	Fork	Spoon	5.25	14	5.88	6	44	61
9	Man	Woman	6.14	1207	5.68	224	66	60
10	Table	Chair	6.00	198	6.12	66	76	31
11	Tale	Story	3.39	21	4.18	153	59	11
12	Sweet	Sour	4.53	70	4.54	3	37	41
13	Rose	Flower	5.86	86	5.62	1	36	25
14	Present	Gift	3.37	377	5.28	33	31	61
15	West	East	4.03	235	3.59	183	78	89

Note. C.cn.= concreteness value of the cues; C.fr.= written frequency of the cues; T.cn. = concreteness value of the targets; T.fr.= written frequency of the targets; F.A.= forward semantic association (from target to cue); B.A.= backward semantic association (from cue to target). Written frequency (Francis & Kucera) and concreteness values of the words based on the MRC Psycholinguistic Database: Machine Usable Dictionary (*source:* http://www.psych.rl.ac.uk/MRC_Psych_Db_files/mrc2.html).

Consent form

I am, Mehmet Akif Guzel, a PhD student in the School of Psychology of the University of Southampton. I am requesting your participation in a study investigating memory and metamemory performance. This will involve the presentations of a list of word pairs and the requirement to remember and report the words form the list. The study will last between 25 and 35 minutes. Personal information will not be released to or viewed by anyone other than researchers involved in this project. Results of the study will not include your name or any other identifying characteristics.

Your participation is voluntary and you may withdraw your participation at any time. If you choose not to participate there will be no consequence to your grade or to your treatment as a student in the psychology department. If you have any questions please ask them now, or contact me, Mehmet Akif Guzel, via the e-mail address, **mag4v07@soton.ac.uk.**

Signature Date

Name Mehmet Akif Guzel

I_________________________________have read the above informed consent form. I understand that I may withdraw my consent and discontinue participation at any time without penalty or loss of benefit to myself. I understand data collected as part of this research project will be treated confidentially, and that published results of this research project will maintain my confidentiality. In signing this consent letter, I am not waiving my legal claims, rights, or remedies. A copy of this letter will be offered to me.

Circle Yes or No

I give consent to participate in the above study. Yes No

Signature Date

Name

I understand that if I have questions about my rights as a participant in this research, or if I feel that I have been placed at risk, I can contact the Chair of the Ethics Committee, Department of Psychology, University of Southampton, Southampton, SO17 1BJ. Phone: (023) 8059 5578.

Reporting sheets for the experimental group

Before you start, please fill in the sections below.

Your age : ______

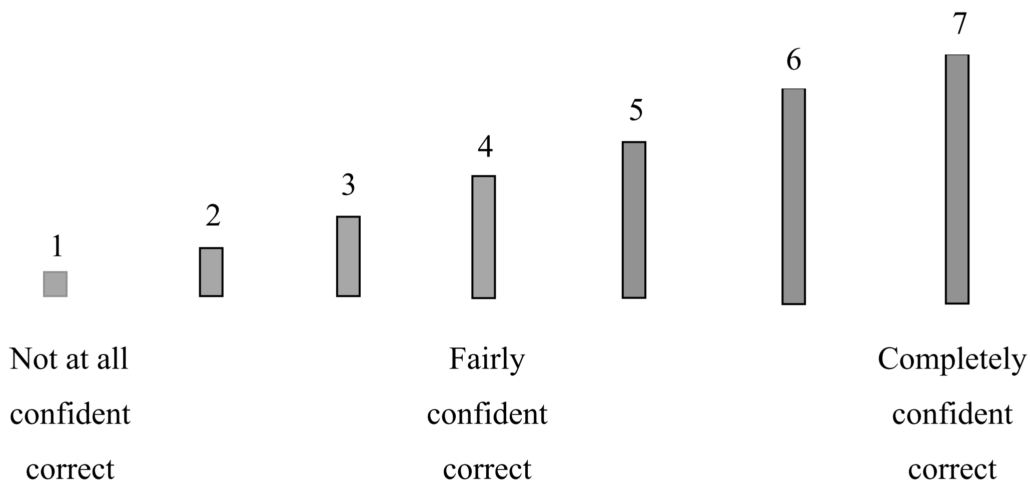
Your gender : Male ____

Female ____

INSTRUCTION

You have been presented a list of word pairs on the computer screen. Now, you are required to report the target words you remember from the list. The target words that you are required to report are the "*WORDS ON THE* ***LEFT*** *SIDE IN THE PAIRS.*"

In order to report, please use the following pages and write down to the target words you remember into the 'TARGETS' column. The target words you remember from the presented list could be reported in any order. That is, reporting order can be made regardless of the order of the list presentation. After you write down each word, please also choose the appropriate number between 1 and 7 provided next to each word according to your confidence level on how well you are sure that your decision is correct. For this, please use the 'CONFIDENCE' column. The numbers in the scale refer to various confidence levels indicated as follows:



TARGETS	CONFIDENCE						
……………………………………	1	2	3	4	5	6	7
……………………………………	1	2	3	4	5	6	7
……………………………………	1	2	3	4	5	6	7
……………………………………	1	2	3	4	5	6	7
……………………………………	1	2	3	4	5	6	7
……………………………………	1	2	3	4	5	6	7
……………………………………	1	2	3	4	5	6	7
……………………………………	1	2	3	4	5	6	7
……………………………………	1	2	3	4	5	6	7
……………………………………	1	2	3	4	5	6	7
……………………………………	1	2	3	4	5	6	7
……………………………………	1	2	3	4	5	6	7
……………………………………	1	2	3	4	5	6	7
……………………………………	1	2	3	4	5	6	7
……………………………………	1	2	3	4	5	6	7
……………………………………	1	2	3	4	5	6	7
……………………………………	1	2	3	4	5	6	7
……………………………………	1	2	3	4	5	6	7
……………………………………	1	2	3	4	5	6	7
……………………………………	1	2	3	4	5	6	7
……………………………………	1	2	3	4	5	6	7
……………………………………	1	2	3	4	5	6	7
……………………………………	1	2	3	4	5	6	7
……………………………………	1	2	3	4	5	6	7
……………………………………	1	2	3	4	5	6	7
……………………………………	1	2	3	4	5	6	7
……………………………………	1	2	3	4	5	6	7
……………………………………	1	2	3	4	5	6	7
……………………………………	1	2	3	4	5	6	7
……………………………………	1	2	3	4	5	6	7

Reporting sheets for the control group

Before you start, please fill in the sections below.

Your age : ______

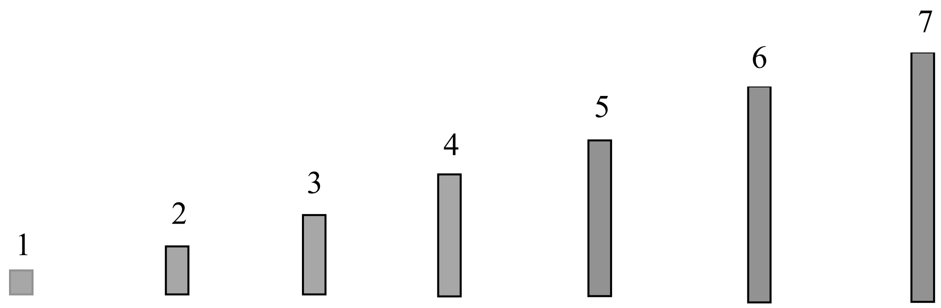
Your gender : Male ____

Female ____

INSTRUCTION

You have been presented a list of word pairs on the computer screen. Now, you are required to report "**as many words as you can**" from the list you have been presented.

In order to report, please use the tables on the following pages and write down to the target words you remember into the 'WORDS' column. The target words you remember from the presented list could be reported in any order. That is, reporting order can be made regardless of the order of the list presentation. After you write down each word, please also choose the appropriate number between 1 and 7 provided next to each word according to your confidence level on how well you are sure that your decision is correct. For this, please use the 'CONFIDENCE' column. The numbers in the scale refer to various confidence levels indicated as follows:



Not at all confident correct	Fairly confident correct	Completely confident correct

WORDS	CONFIDENCE						
…………………………………………	1	2	3	4	5	6	7
…………………………………………	1	2	3	4	5	6	7
…………………………………………	1	2	3	4	5	6	7
…………………………………………	1	2	3	4	5	6	7
…………………………………………	1	2	3	4	5	6	7
…………………………………………	1	2	3	4	5	6	7
…………………………………………	1	2	3	4	5	6	7
…………………………………………	1	2	3	4	5	6	7
…………………………………………	1	2	3	4	5	6	7
…………………………………………	1	2	3	4	5	6	7
…………………………………………	1	2	3	4	5	6	7
…………………………………………	1	2	3	4	5	6	7
…………………………………………	1	2	3	4	5	6	7
…………………………………………	1	2	3	4	5	6	7
…………………………………………	1	2	3	4	5	6	7
…………………………………………	1	2	3	4	5	6	7
…………………………………………	1	2	3	4	5	6	7
…………………………………………	1	2	3	4	5	6	7
…………………………………………	1	2	3	4	5	6	7
…………………………………………	1	2	3	4	5	6	7
…………………………………………	1	2	3	4	5	6	7
…………………………………………	1	2	3	4	5	6	7
…………………………………………	1	2	3	4	5	6	7
…………………………………………	1	2	3	4	5	6	7
…………………………………………	1	2	3	4	5	6	7
…………………………………………	1	2	3	4	5	6	7
…………………………………………	1	2	3	4	5	6	7
…………………………………………	1	2	3	4	5	6	7
…………………………………………	1	2	3	4	5	6	7
…………………………………………	1	2	3	4	5	6	7
…………………………………………	1	2	3	4	5	6	7
…………………………………………	1	2	3	4	5	6	7
…………………………………………	1	2	3	4	5	6	7
…………………………………………	1	2	3	4	5	6	7

……………………………………	1	2	3	4	5	6	7
……………………………………	1	2	3	4	5	6	7
……………………………………	1	2	3	4	5	6	7
……………………………………	1	2	3	4	5	6	7
……………………………………	1	2	3	4	5	6	7
……………………………………	1	2	3	4	5	6	7
……………………………………	1	2	3	4	5	6	7
……………………………………	1	2	3	4	5	6	7
……………………………………	1	2	3	4	5	6	7
……………………………………	1	2	3	4	5	6	7
……………………………………	1	2	3	4	5	6	7
……………………………………	1	2	3	4	5	6	7
……………………………………	1	2	3	4	5	6	7
……………………………………	1	2	3	4	5	6	7
……………………………………	1	2	3	4	5	6	7
……………………………………	1	2	3	4	5	6	7
……………………………………	1	2	3	4	5	6	7
……………………………………	1	2	3	4	5	6	7
……………………………………	1	2	3	4	5	6	7
……………………………………	1	2	3	4	5	6	7
……………………………………	1	2	3	4	5	6	7
……………………………………	1	2	3	4	5	6	7
……………………………………	1	2	3	4	5	6	7
……………………………………	1	2	3	4	5	6	7
……………………………………	1	2	3	4	5	6	7
……………………………………	1	2	3	4	5	6	7
……………………………………	1	2	3	4	5	6	7
……………………………………	1	2	3	4	5	6	7
……………………………………	1	2	3	4	5	6	7
……………………………………	1	2	3	4	5	6	7
……………………………………	1	2	3	4	5	6	7
……………………………………	1	2	3	4	5	6	7
……………………………………	1	2	3	4	5	6	7
……………………………………	1	2	3	4	5	6	7
……………………………………	1	2	3	4	5	6	7

…………………………………	1	2	3	4	5	6	7
…………………………………	1	2	3	4	5	6	7

Debriefing form

The aim of the research was to investigate the effect of the semantic association between word pairs on the memory and metamemory performance. It is expected that the participants would recall more target words in terms of number but would be less accurate in their decisions as a response to the target words paired with their high associates than paired with low associates. Your data will help our understanding of how the memory and metamemory performance are affected by the semantic association level. The study did not use any deception. You may have a copy of this summary and if you wish you may also learn the summary of research findings with the following contact information.

If you have any further questions please contact me, Mehmet Akif Guzel, by the following e-mail address: mag4v07@soton.ac.uk.
Thank you for your participation in this research.

Signature ______________________________ Date __________________
Name: Mehmet Akif Guzel

I understand that if I have questions about my rights as a participant in this research, or if I feel that I have been placed at risk, I can contact the Chair of the Ethics Committee, Department of Psychology, University of Southampton, Southampton, SO17 1BJ. Phone: (023) 8059 5578.

Appendix B

Materials used in Experiment 2

Table 19

Study Lists Having High and Low Inter-Target Associations (ITA)

#	CUES	TARGETS	C.cn	C.fr.	T.cn.	T.fr.	NORM
			High-ITA List-1				
1	School	Steel	5.25	492	0.00	9	62
2	Victim	Iron	5.49	27	5.47	43	57
3	Tennis	Silver	5.54	15	5.36	29	55
4	Movie	Copper	5.85	29	4.98	13	53
5	Street	Gold	5.84	244	5.68	52	53
6	Shower	Dog	5.88	15	5.75	75	98
7	People	Cat	5.51	847	6.21	23	97
8	Towel	Horse	6.83	6	6.03	117	52
9	Drum	Lion	-	-	5.81	57	41
10	Calendar	Bear	6.48	28	5.81	57	37
11	Plane	Blue	5.31	114	4.49	143	100
12	Garbage	Red	5.68	7	4.97	197	96
13	Doctor	Green	5.75	100	5.46	116	93
14	Car	Yellow	6.35	274	5.18	55	92
15	Music	Black	5.15	216	4.66	203	77
16	Bottle	Chair	6.94	76	6.12	66	90
17	Star	Table	5.59	25	6.00	198	75
18	Muscle	Couch	6.00	42	5.74	12	70
19	Doll	Bed	5.36	10	5.15	127	58
20	Cake	Desk	6.11	13	5.79	65	49
			High-ITA List-2				
21	Circus	Shirt	5.31	7	6.05	27	90
22	Ocean	Pants	5.63	34	6.15	9	85
23	Puppet	Socks	-	-	5.46	7	76
24	Honey	Shoe	5.91	25	5.96	14	57
25	Vase	Hat	5.70	4	5.97	56	53
26	Mall	Beer	5.94	3	5.83	34	87
27	Skin	Vodka	6.96	47	-	-	62
28	Lake	Wine	5.70	54	6.40	72	54
29	Perfume	Rum	6.03	10	5.96	3	43
30	Agenda	Whiskey	0.00	5	6.00	17	32
31	Island	Priest	6.40	167	6.59	16	71
32	Razor	Pope	6.32	15	-	-	52
33	Nail	Bishop	5.60	6	5.83	18	33
34	Bike	Nun	4.61	0	6.76	2	29
35	Mask	Cardinal	6.38	9	5.79	16	20
36	Dice	Apple	0.00	14	7.00	9	95
37	Baby	Banana	5.77	62	6.29	4	71
38	Chess	Grape	0.00	3	5.85	3	52
39	Tobacco	Pear	6.16	19	6.30	6	50
40	Lens	Peach	5.52	12	6.05	3	40

Table 19 (*continued*).

#	CUES	TARGETS	C.cn	C.fr.	T.cn.	T.fr.	NORM
			Low-ITA List-1				
1	School	Door	5.25	492	5.95	312	-
2	Victim	Dollars	5.49	27	5.68	46	-
3	Tennis	Mirror	5.54	15	5.91	157	-
4	Movie	Paper	5.85	29	5.96	157	-
5	Street	Shampoo	5.84	244	0.00	2	-
6	Shower	File	5.88	15	4.63	81	-
7	People	Rock	5.51	847	6.03	75	-
8	Towel	Hammer	6.83	6	5.77	9	-
9	Drum	Bulb	-	-	6.12	0	-
10	Calendar	Picture	6.48	28	6.75	162	-
11	Plane	Tunnel	5.31	114	5.51	10	-
12	Garbage	Farm	5.68	7	5.53	125	-
13	Doctor	Candle	5.75	100	5.37	18	-
14	Car	Office	6.35	274	5.65	255	-
15	Music	Dirt	5.15	216	5.51	43	-
16	Bottle	King	6.94	76	-	-	-
17	Star	Brick	5.59	25	5.97	18	-
18	Muscle	Phone	6.00	42	6.02	54	-
19	Doll	Friend	5.36	10	4.40	133	-
20	Cake	Summer	6.11	13	4.68	134	-
			Low-ITA List-2				
21	Circus	Army	5.31	7	6.53	132	-
22	Ocean	Coffee	563	34	6.43	78	-
23	Puppet	Girl	-	-	6.83	220	-
24	Honey	Computer	5.91	25	0.00	7	-
25	Vase	Clock	5.70	4	6.94	20	-
26	Mall	Volcano	5.94	3	6.83	2	-
27	Skin	Wedding	6.96	47	5.54	32	-
28	Lake	Picnic	5.70	54	5.40	15	-
29	Perfume	Fire	6.03	10	6.13	187	-
30	Agenda	Flag	0.00	5	6.20	16	-
31	Island	Curtain	6.40	167	6.82	13	-
32	Razor	Bone	6.32	15	5.75	33	-
33	Nail	Laundry	5.60	6	6.10	5	-
34	Bike	Coin	4.61	0	5.70	10	-
35	Mask	Toast	6.38	9	5.54	19	-
36	Dice	Atom	0.00	14	4.77	37	-
37	Baby	Castle	5.77	62	6.50	7	-
38	Chess	Wallet	0.00	3	5.71	6	-
39	Tobacco	Soil	6.16	19	5.70	54	-
40	Lens	Balloon	5.52	12	6.10	10	-

Note. C.cn.= concreteness value of the cues; C.fr.= written frequency of the cues; T.cn. = concreteness value of the targets; T.fr.= written frequency of the targets; NORM = free association norms which display the percentage (%) of 642 participants who gave the response as an exemplar of the particular category asked in the norms study of Van Overschelde, Rawson, and Dunlosky (2004). Written frequency (Francis & Kucera) and concreteness values of the words based on the MRC Psycholinguistic Database: Machine Usable Dictionary (*source:* http://www.psych.rl.ac.uk/MRC_Psych_Db_files/mrc2.html).

The eight categories (four of which is in high-ITA List-1 and four of which is in high-ITA List-2) were matched in terms of their mean exemplar dominancy. Four pairs of categories were investigated with t-tests as to whether the matched pairs differ from each other in terms of mean exemplar dominancy. Amongst four category pairs, only two of the categories (*colours* in List-1 vs. *articles of clothing* in List-2) differed from each other (M = 91.6, SD = 8.73; M = 69.6, SD = 20.15, respectively), $t(4) = 3.719$, p = .02. The remaining three category pairs matched in terms of dominancy and which were: *colour*s (M = 56.0, SD = 3.74) vs. *member of clergy* (M = 41.0, SD = 20.43); *animals* (M = 66.00, SD = 29.72) vs. *alcoholic beverages* (M = 55.60, SD = 20.89); and, *furniture items* (M = 68.4, SD = 15.79) vs. *fruits* (M = 61.60, SD = 21.78) did not differ from each other in terms of mean exemplar dominancy; $t(4) = 1.996$, $p > .05$; $t(4) = 1.602$, $p > .05$, and $t(4) = 1.821$, $p > .05$, respectively.

One might wonder whether the written frequency values of the target words might have been a factor in yielding greater memory performance in high-ITA target recall compared to low-ITA target recall. However, the written frequency means were 82.9 (SD = 63.2), 17.6 (SD = 19.20), 89.6 (SD = 88.5), and 45.15 (SD =62.5) for High-ITA List-1, High-ITA List-2, Low-ITA List-1, and Low-ITA List-2, respectively. T-test mean comparisons showed that High-ITA List-1 and Low-ITA List-1 did not differ in terms of written frequency, $t(19) = -.863$, $p > .05$, and High-ITA List-2 and Low-ITA List-2 did not differ in term of written frequency means as well, $t(17) = -1.804$, $p = .09$. In other words, greater memory performance observed in High-ITA target recall compared to Low-ITA target recall did not seem to be resulted from written frequency values.

Consent Form

I am Mehmet Akif Guzel a PhD student in the School of Psychology of the University of Southampton. I am requesting your participation in a study regarding the effects of contextual variations on memory and metamemory performances. This will involve presentations of a list of word pairs and the requirement to remember and report the words form the list. The study will last between 30 and 40 minutes. Personal information will not be released to or viewed by anyone other than researchers involved in this project. Results of the study will not include your name or any other identifying characteristics.

Your participation is voluntary and you may withdraw your participation at any time. If you choose not to participate there will be no consequence to your grade or to your treatment as a student in the psychology department. If you have any questions please ask them now, or contact me, Mehmet Akif Guzel, via the e-mail address, mag4v07@soton.ac.uk.

Signature Date

Name Mehmet Akif Guzel

I__________________________________have read the above informed consent form. I understand that I may withdraw my consent and discontinue participation at any time without penalty or loss of benefit to myself. I understand data collected as part of this research project will be treated confidentially, and that published results of this research project will maintain my confidentiality. In signing this consent letter, I am not waiving my legal claims, rights, or remedies. A copy of this letter will be offered to me.

Circle Yes or No

I give consent to participate in the above study. Yes No

Signature Date

Name

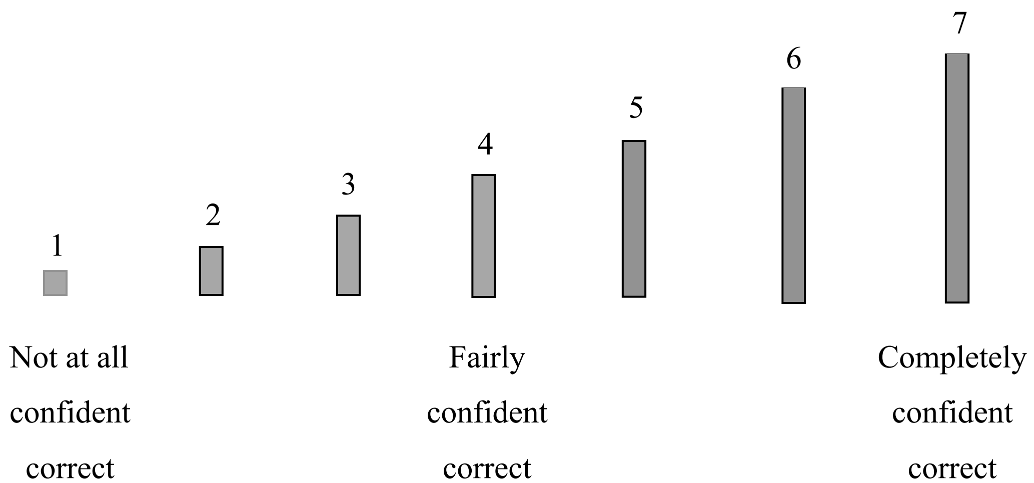
I understand that if I have questions about my rights as a participant in this research, or if I feel that I have been placed at risk, I can contact the Chair of the Ethics Committee, Department of Psychology, University of Southampton, Southampton, SO17 1BJ. Phone: (023) 8059 5578.

Reporting sheets for the free-report cued-recall group to be used for the 1st list

INSTRUCTIONS

You have been presented a list of word pairs on the computer screen. Now, you are required to report the target words you remember from the list. The target words that you are required to report are the "*WORDS THAT APPEARED ON THE* ***RIGHT*** *SIDE IN THE PAIRS.*" Use the "*CUE WORDS*", the words that appeared on the left-hand-side in the pairs, to help you remember the targets. These are shown on the next page.

For reporting, please use the following page and write down the target words you remember into the "TARGETS" column. Please try to report the word that you think was paired with the cue provided. After you write down each word, please also choose the appropriate number between 1 and 7 provided next to each word to indicate how confident you are that your decision is correct. To do this, please use the "CONFIDENCE" column. The numbers in the scale refer to various confidence levels in the following manner:



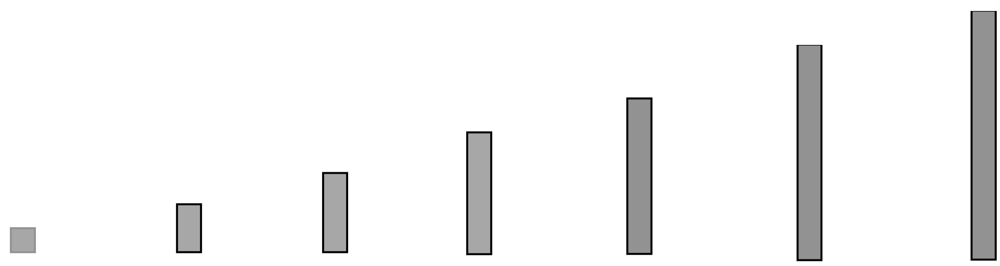
Also, do not forget you have no time limitation for reporting!!

Reporting sheets for the forced-report cued-recall group to be used for the 1st list

INSTRUCTIONS

You have been presented a list of word pairs on the computer screen. Now, you are required to report the target words you remember from the list. The target words that you are required to report are the "*WORDS THAT APPEARED ON THE **RIGHT** SIDE IN THE PAIRS.*" Use the "*CUE WORDS*", the words that appeared on the left-hand-side in the pairs, to help you remember the targets. These are shown on the next page.

For reporting, please use the following page and write down the target words you remember into the "TARGETS" column. Please try to report the word that you think was paired with the cue provided. After you write down each word, please also choose the appropriate number between 1 and 7 provided next to each word to indicate how confident you are that your decision is correct. To do this, please use the "CONFIDENCE" column. The numbers in the scale refer to various confidence levels in the following manner:



Not at all confident correct — Fairly confident correct — Completely confident correct

IMPORTANT: YOU ARE REQUIRED TO FILL <u>ALL OF THE EMPTY SPACES IN THE TARGETS COLUMN</u> WITH TARGET WORDS THAT YOU REMEMBER. EVEN IF YOU FEEL YOU CANNOT REMEMBER ANY MORE TARGETS, PLEASE MAKE YOUR BEST GUESS. YOU NEVER KNOW; YOU MIGHT BE RIGHT! IF YOU HAVE TO GUESS, INDICATE THIS WITH THE CONFIDENCE LEVEL 1.

Also, do not forget you have no time limitation for reporting!!

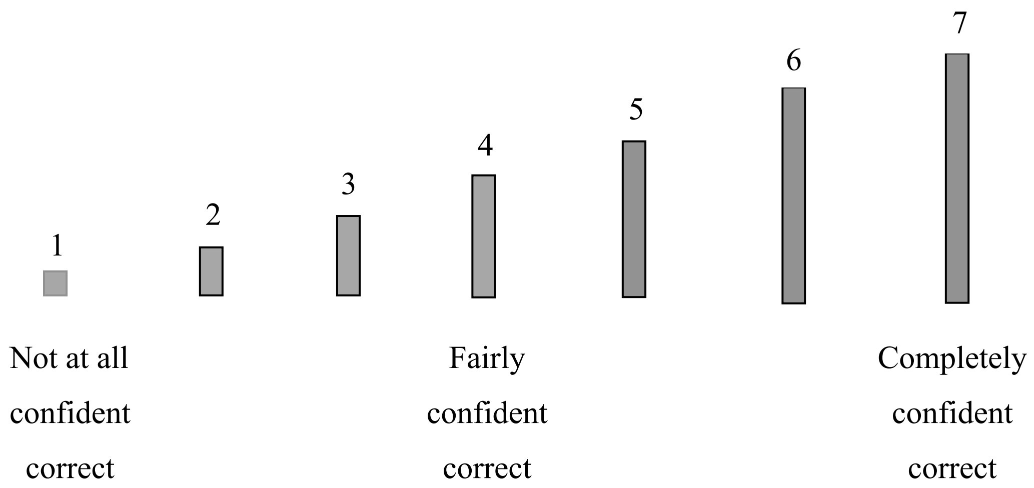
CUES	TARGETS	CONFIDENCE						
DOLL	…………………….	1	2	3	4	5	6	7
STREET	…………………….	1	2	3	4	5	6	7
PEOPLE	…………………….	1	2	3	4	5	6	7
DOCTOR	…………………….	1	2	3	4	5	6	7
SCHOOL	…………………….	1	2	3	4	5	6	7
TENNIS	…………………….	1	2	3	4	5	6	7
GARBAGE	…………………….	1	2	3	4	5	6	7
BOTTLE	…………………….	1	2	3	4	5	6	7
STAR	…………………….	1	2	3	4	5	6	7
SHOWER	…………………….	1	2	3	4	5	6	7
CAR	…………………….	1	2	3	4	5	6	7
CAKE	…………………….	1	2	3	4	5	6	7
MUSCLE	…………………….	1	2	3	4	5	6	7
TOWEL	…………………….	1	2	3	4	5	6	7
MUSIC	…………………….	1	2	3	4	5	6	7
VICTIM	…………………….	1	2	3	4	5	6	7
DRUM	…………………….	1	2	3	4	5	6	7
PLANE	…………………….	1	2	3	4	5	6	7
MOVIE	…………………….	1	2	3	4	5	6	7
CALENDAR	…………………….	1	2	3	4	5	6	7

Reporting sheets for the cued-recall free-report group to be used for the 2nd list

INSTRUCTIONS

You have been presented a second list of word pairs on the computer screen. Now, you are also required to report the target words you remember from the list. The target words now are the "*WORDS ON THE **RIGHT** SIDE IN THE PAIRS*" presented in the *SECOND LIST*. Use the "*CUE WORDS*", the words that appeared on the left-hand-side in the pairs, to help you remember the targets. These are shown on the next page.

For reporting, please use the TARGETS column on the next page and indicate your confidence level using the following scale:



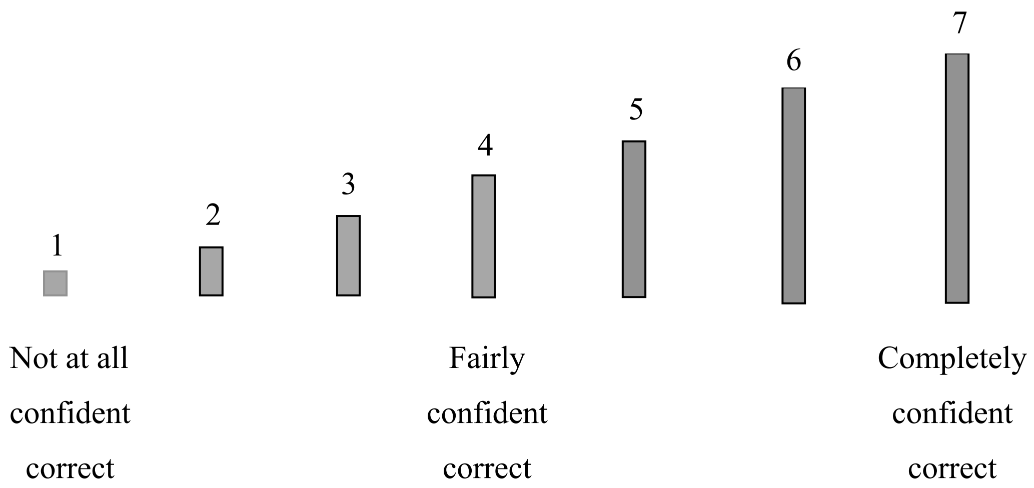
Also, do not forget you do not have time limitation for reporting!

Reporting sheets for the forced-report cued-recall group to be used for the 2nd list

INSTRUCTIONS

You have been presented a second list of word pairs on the computer screen. Now, you are also required to report the target words you remember from the list. The target words now are the "*WORDS ON THE **RIGHT** SIDE IN THE PAIRS*" presented in the *SECOND LIST*. Use the "*CUE WORDS*", the words that appeared on the left-hand-side in the pairs, to help you remember the targets. These are shown on the next page.

For reporting, please use the TARGETS column on the next page and indicate your confidence level using the following scale:



IMPORTANT: YOU ARE REQUIRED TO FILL <u>ALL OF THE EMPTY SPACES IN THE TARGETS COLUMN</u> WITH TARGET WORDS THAT YOU REMEMBER. EVEN IF YOU FEEL YOU CANNOT REMEMBER ANY MORE TARGETS, PLEASE MAKE YOUR BEST GUESS. YOU NEVER KNOW; YOU MIGHT BE RIGHT! IF YOU HAVE TO GUESS, INDICATE THIS WITH THE CONFIDENCE LEVEL 1.

Also, do not forget you do not have time limitation for reporting!

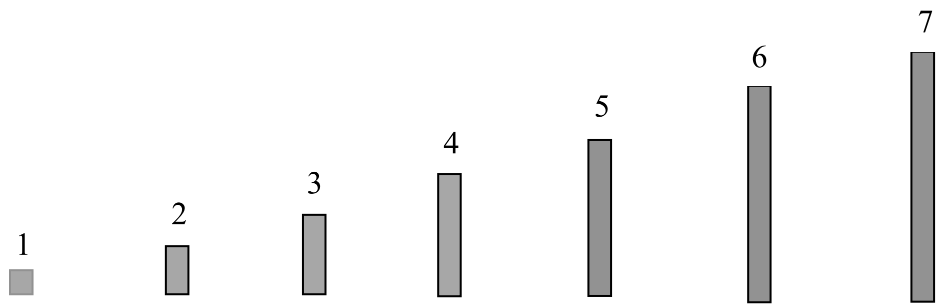
CUES	TARGETS	CONFIDENCE						
PERFUME	…………………….	1	2	3	4	5	6	7
DICE	…………………….	1	2	3	4	5	6	7
CIRCUS	…………………….	1	2	3	4	5	6	7
SKIN	…………………….	1	2	3	4	5	6	7
LENS	…………………….	1	2	3	4	5	6	7
NAIL	…………………….	1	2	3	4	5	6	7
VASE	…………………….	1	2	3	4	5	6	7
BIKE	…………………….	1	2	3	4	5	6	7
BABY	…………………….	1	2	3	4	5	6	7
LAKE	…………………….	1	2	3	4	5	6	7
MALL	…………………….	1	2	3	4	5	6	7
OCEAN	…………………….	1	2	3	4	5	6	7
MASK	…………………….	1	2	3	4	5	6	7
CHESS	…………………….	1	2	3	4	5	6	7
PUPPET	…………………….	1	2	3	4	5	6	7
ISLAND	…………………….	1	2	3	4	5	6	7
AGENDA	…………………….	1	2	3	4	5	6	7
TOBACCO	…………………….	1	2	3	4	5	6	7
HONEY	…………………….	1	2	3	4	5	6	7
RAZOR	…………………….	1	2	3	4	5	6	7

Reporting sheets for the free-report uncued-recall group to be used for the 1st list

INSTRUCTIONS

You have been presented a list of word pairs on the computer screen. Now, you are required to report the target words you remember from the list. The target words that you are required to report are the "*WORDS ON THE **RIGHT** SIDE IN THE PAIRS*".

For reporting, please use the table on the following page and write down the target words you remember into the "TARGETS" column. After you write down each word, please also choose the appropriate number between 1 and 7 provided next to each word to indicate how confident you are that your decision is correct. To do this, please use the "CONFIDENCE" column. The numbers in the scale refer to various confidence levels in the following manner:



Not at all confident correct	Fairly confident correct	Completely confident correct

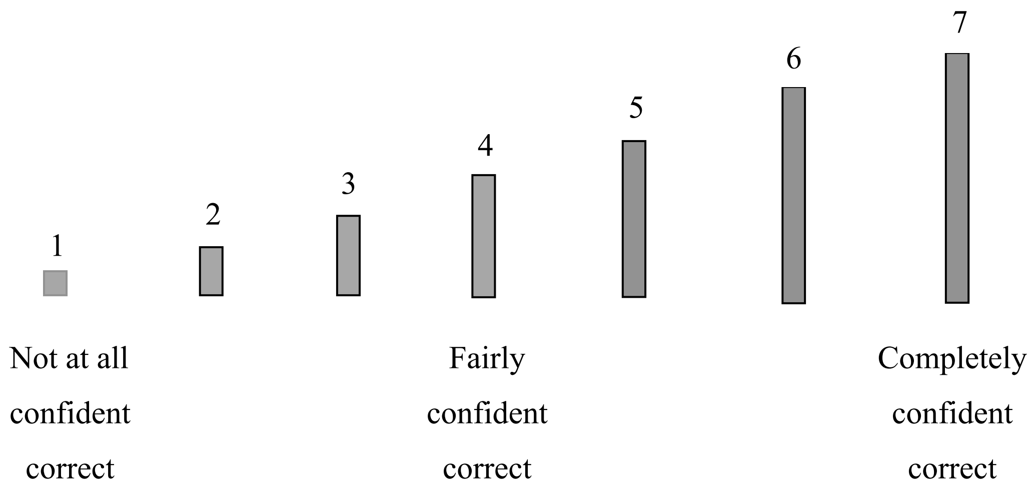
Also do not forget you have no time limitation for reporting!!

Reporting sheets for the forced-report uncued-recall group to be used for the 1st list

INSTRUCTIONS

You have been presented a list of word pairs on the computer screen. Now, you are required to report the target words you remember from the list. The target words that you are required to report are the "*WORDS ON THE **RIGHT** SIDE IN THE PAIRS*".

For reporting, please use the table on the following page and write down the target words you remember into the "TARGETS" column. After you write down each word, please also choose the appropriate number between 1 and 7 provided next to each word to indicate how confident you are that your decision is correct. To do this, please use the "CONFIDENCE" column. The numbers in the scale refer to various confidence levels in the following manner:



IMPORTANT: YOU ARE REQUIRED TO FILL ALL OF THE EMPTY SPACES IN THE TARGETS COLUMN WITH TARGET WORDS THAT YOU REMEMBER. EVEN IF YOU FEEL YOU CANNOT REMEMBER ANY MORE TARGETS, PLEASE MAKE YOUR BEST GUESS. YOU NEVER KNOW; YOU MIGHT BE RIGHT! IF YOU HAVE TO GUESS, INDICATE THIS WITH THE CONFIDENCE LEVEL 1.

Also do not forget you have no time limitation for reporting!!

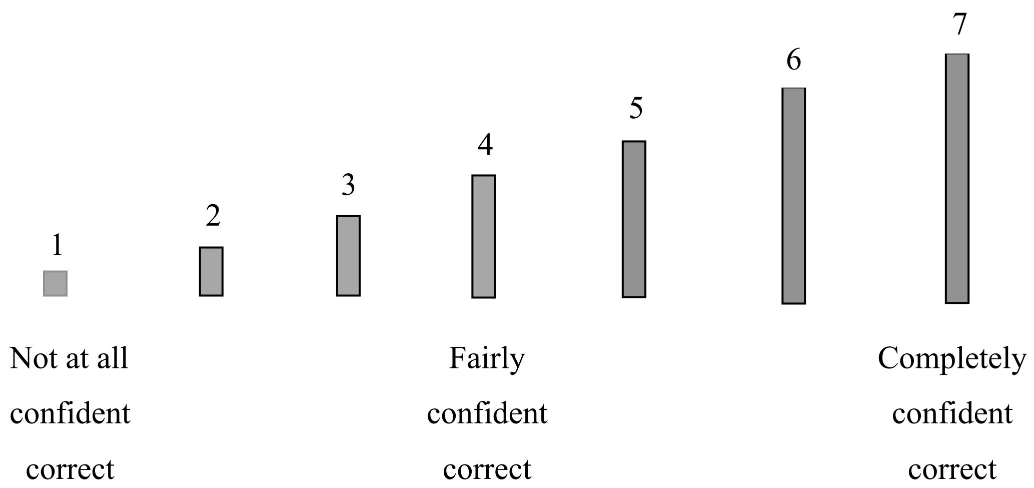
TARGETS	CONFIDENCE						
…………………………….	1	2	3	4	5	6	7
…………………………….	1	2	3	4	5	6	7
…………………………….	1	2	3	4	5	6	7
…………………………….	1	2	3	4	5	6	7
…………………………….	1	2	3	4	5	6	7
…………………………….	1	2	3	4	5	6	7
…………………………….	1	2	3	4	5	6	7
…………………………….	1	2	3	4	5	6	7
…………………………….	1	2	3	4	5	6	7
…………………………….	1	2	3	4	5	6	7
…………………………….	1	2	3	4	5	6	7
…………………………….	1	2	3	4	5	6	7
…………………………….	1	2	3	4	5	6	7
…………………………….	1	2	3	4	5	6	7
…………………………….	1	2	3	4	5	6	7
…………………………….	1	2	3	4	5	6	7
…………………………….	1	2	3	4	5	6	7
…………………………….	1	2	3	4	5	6	7
…………………………….	1	2	3	4	5	6	7
…………………………….	1	2	3	4	5	6	7

Reporting sheets for the free-report uncued-recall group to be used for the 2nd list

INSTRUCTION

You have been presented a second list of word pairs on the computer screen. Now, you are required to report the target words that are the "*WORDS ON THE **RIGHT** SIDE IN THE PAIRS*" presented in the *SECOND LIST*.

For reporting, please use the TARGETS column on the next page and indicate your confidence level using the following scale:



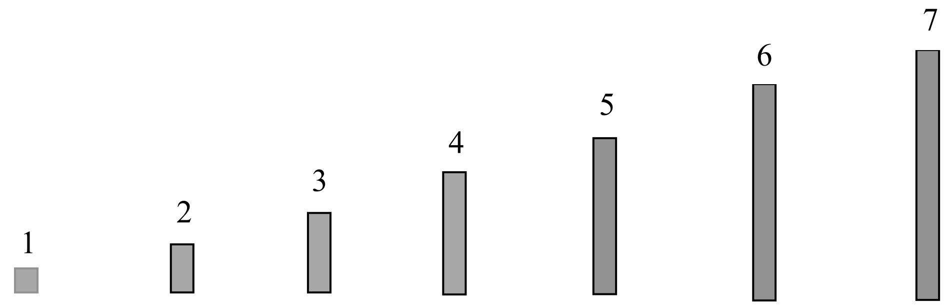
Also do not forget you have no time limitation for reporting!!

Reporting sheets for the forced-report uncued-recall group to be used for the 2nd list

INSTRUCTION

You have been presented a second list of word pairs on the computer screen. Now, you are required to report the target words that are the "*WORDS ON THE **RIGHT** SIDE IN THE PAIRS*" presented in the *SECOND LIST*.

For reporting, please use the TARGETS column on the next page and indicate your confidence level using the following scale:



Not at all confident correct — Fairly confident correct — Completely confident correct

IMPORTANT: YOU ARE REQUIRED TO FILL ALL OF THE EMPTY SPACES IN THE TARGETS COLUMN WITH TARGET WORDS THAT YOU REMEMBER. EVEN IF YOU FEEL YOU CANNOT REMEMBER ANY MORE TARGETS, PLEASE MAKE YOUR BEST GUESS. YOU NEVER KNOW; YOU MIGHT BE RIGHT! IF YOU HAVE TO GUESS, INDICATE THIS WITH THE CONFIDENCE LEVEL 1.

Also do not forget you have no time limitation for reporting!!

TARGETS				**CONFIDENCE**			
……………………………	1	2	3	4	5	6	7
……………………………	1	2	3	4	5	6	7
……………………………	1	2	3	4	5	6	7
……………………………	1	2	3	4	5	6	7
……………………………	1	2	3	4	5	6	7
……………………………	1	2	3	4	5	6	7
……………………………	1	2	3	4	5	6	7
……………………………	1	2	3	4	5	6	7
……………………………	1	2	3	4	5	6	7
……………………………	1	2	3	4	5	6	7
……………………………	1	2	3	4	5	6	7
……………………………	1	2	3	4	5	6	7
……………………………	1	2	3	4	5	6	7
……………………………	1	2	3	4	5	6	7
……………………………	1	2	3	4	5	6	7
……………………………	1	2	3	4	5	6	7
……………………………	1	2	3	4	5	6	7
……………………………	1	2	3	4	5	6	7
……………………………	1	2	3	4	5	6	7
……………………………	1	2	3	4	5	6	7

Debriefing Statement

The aim of the research was to investigate the effect of report option (free vs. forced) and the inter-item association level of the word pairs on memory and metamemory performance. It is expected that the participants would recall more target words in forced-report options compared to free-report options. Also, high inter-target association would yield higher memory performance but comparable metamemory performance compared to low inter-target association lists. Your data will help our understanding of how the memory performances are affected by the inter-item association level as a contextual information. The study did not use any deception. You may have a copy of this summary and if you wish you may also learn the summary of research findings with the following contact information.

If you have any further questions please contact me, Mehmet Akif Guzel, by the following e-mail address: mag4v07@soton.ac.uk.

Thank you for your participation in this research.

Signature _________________________________ Date ___________________

Name

I understand that if I have questions about my rights as a participant in this research, or if I feel that I have been placed at risk, I can contact the Chair of the Ethics Committee, Department of Psychology, University of Southampton, Southampton, SO17 1BJ. Phone: (023) 8059 5578.

Appendix C

Materials used in Experiment 3

The study lists which were utilised in Experiment 2 were also used in Experiment 3; see Appendix B.

Consent Form

I am Mehmet Akif Guzel a PhD student in the School of Psychology of the University of Southampton. I am requesting your participation in a study regarding the effects of contextual variations on memory and metamemory performances. This will involve presentations of a list of word pairs and the requirement to remember and report the words form the list. The study will last between 30 and 40 minutes. Personal information will not be released to or viewed by anyone other than researchers involved in this project. Results of the study will not include your name or any other identifying characteristics.

Your participation is voluntary and you may withdraw your participation at any time. If you choose not to participate there will be no consequence to your grade or to your treatment as a student in the psychology department. If you have any questions please ask them now, or contact me, Mehmet Akif Guzel, via the e-mail address, mag4v07@soton.ac.uk.

Signature Date

Name Mehmet Akif Guzel

I__________________________________have read the above informed consent form. I understand that I may withdraw my consent and discontinue participation at any time without penalty or loss of benefit to myself. I understand data collected as part of this research project will be treated confidentially, and that published results of this research project will maintain my confidentiality. In signing this consent letter, I am not waiving my legal claims, rights, or remedies. A copy of this letter will be offered to me.

Circle Yes or No

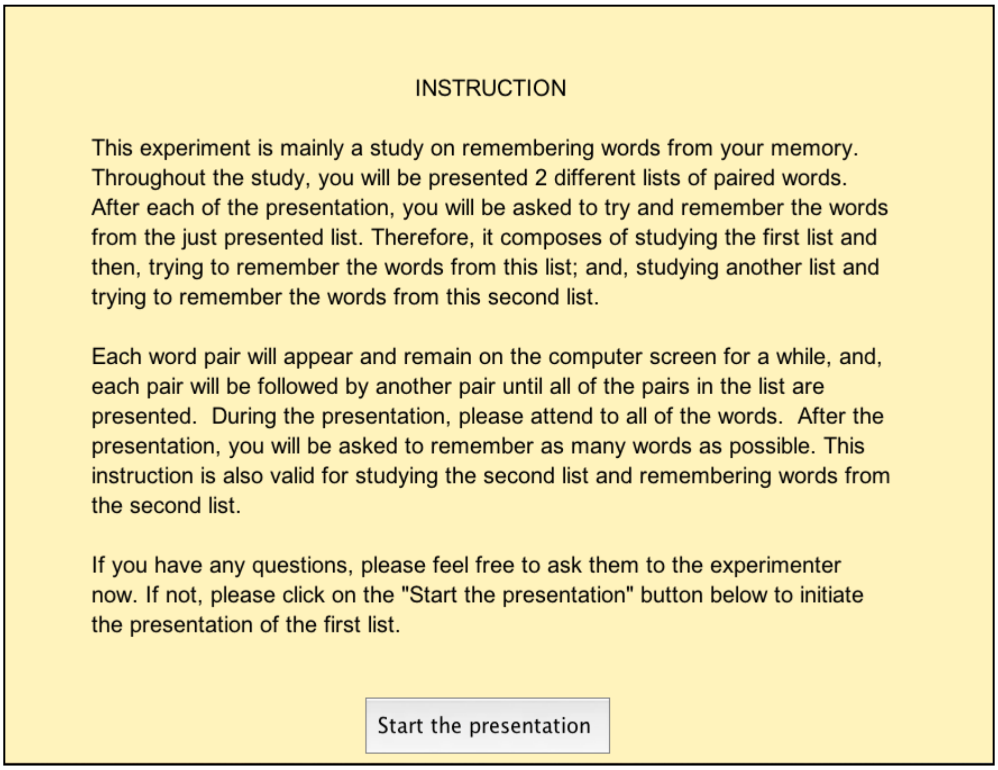
I give consent to participate in the above study. Yes No

Signature Date …../ …../ ……..

Name

I understand that if I have questions about my rights as a participant in this research, or if I feel that I have been placed at risk, I can contact the Chair of the Ethics Committee, Department of Psychology, University of Southampton, Southampton, SO17 1BJ. Phone: (023) 8059 5578.

a) Instruction for the pair-focused way of encoding



b) Instruction for the target-focused way of encoding

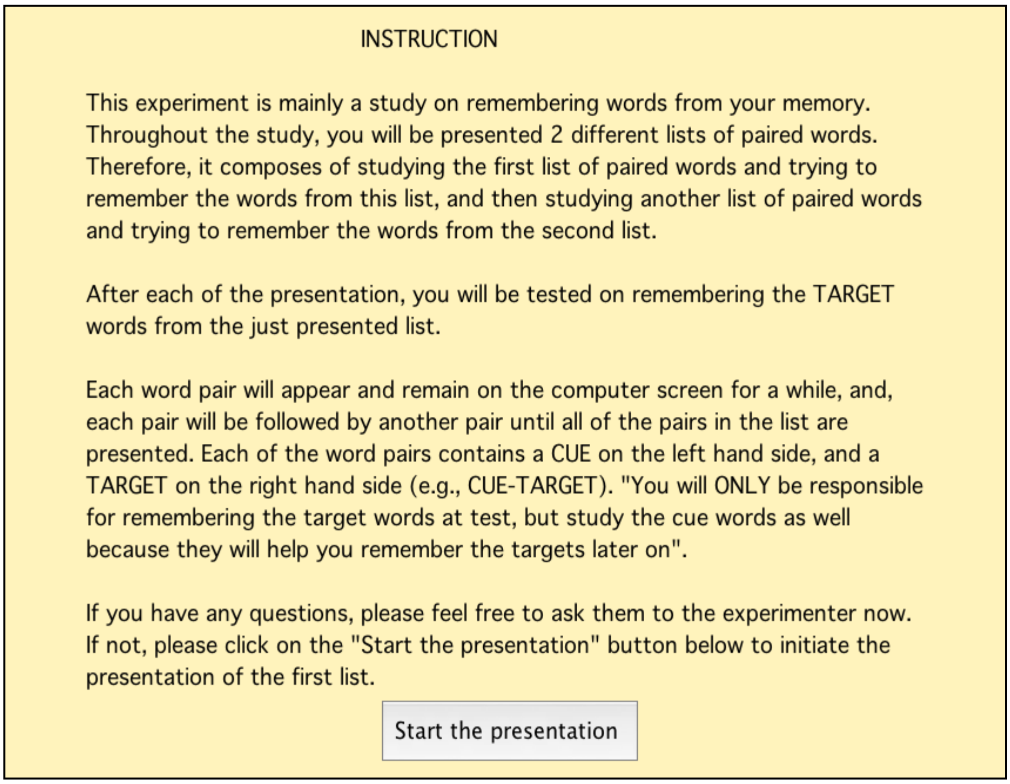


Figure 16. Instructions used for the pair-focused encoding strategy and the target-focused encoding strategy. The pages were taken from the Runtime Revolution computer program used for the study.

Reporting sheets for the constrained cued-recall group to be used for the 1st list

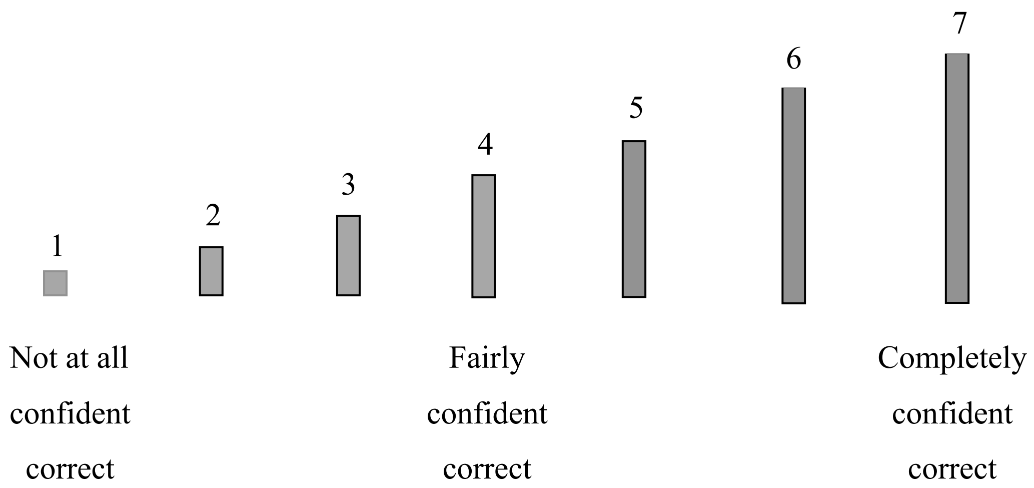
INSTRUCTIONS

You have been presented a list of word pairs on the computer screen. Now, you are asked to report the target words you remember from the list. The target words are the "*WORDS ON THE* ***RIGHT*** *SIDE IN THE PAIRS.*" To remember the words, utilise the "*CUES*, the left hand side words in the pairs studied, given to you on the next pages.

Importantly, you are required to give a response to every single cue, even if you have to guess. Write down your responses in the "**Targets**" column. Sometimes you may prefer to "**pass**" because you can't remember the target that was presented with the cue. If so, still write down your best guess, but check the "**pass**" checkbox next to that item. For all other responses, check the "**report**" checkbox, indicating that you are comfortable giving an answer to that item.

IMPORTANT: A response will be counted correct *only if* the target word you give matches with the cue studied with. In other words, you need to respond to every cue with the target word studied together.

In addition to writing down a response to every single item and checking the appropriate checkbox, also rate your confidence using the following 1-7 scale:



Reporting sheets for liberated cued-recall group to be used for the 2nd list

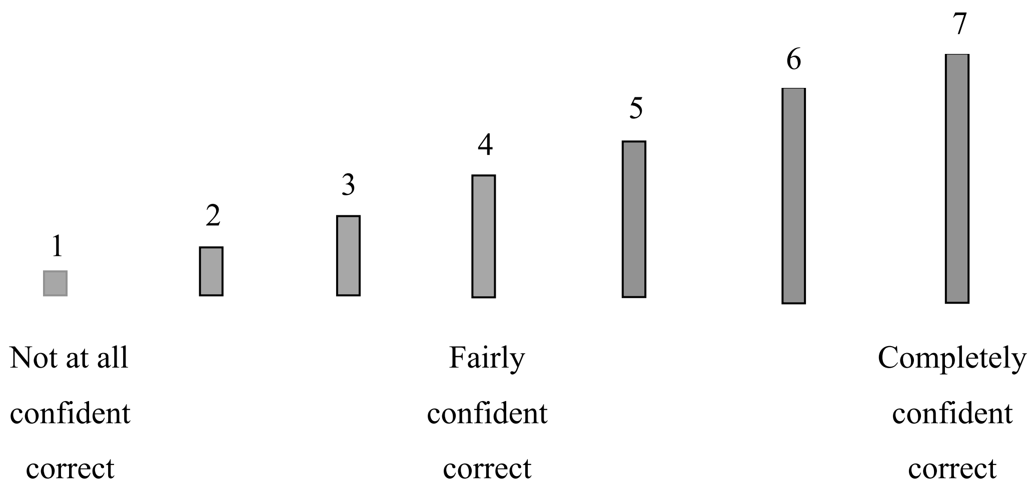
INSTRUCTIONS

You have been presented a second list of word pairs on the computer screen. Now, you are also asked to report the target words you remember from the list. The target words are the "*WORDS ON THE* ***RIGHT*** *SIDE IN THE PAIRS.*" To remember the words from the second list, again use the "*CUES*" given to you on the next pages.

Importantly, you are again required to give a response to every single cue, even if you have to guess. Write down your responses in the "**Targets**" column. Sometimes you may prefer to "**pass**" because you can't remember the target that was presented with the cue. If so, still write down your best guess, but check the "**pass**" checkbox next to that item. For all other responses, check the "**report**" checkbox, indicating that you are comfortable giving an answer to that item.

IMPORTANT: Again, your responses will be counted correct *only if* the target word you give matches with the cue studied with. In other words, you need to respond to every cue with the target word studied together.

In addition to writing down a response to every single item and checking the appropriate checkbox, also rate your confidence using the following 1-7 scale:



Reporting sheets for liberated cued-recall group to be used for the 1st list

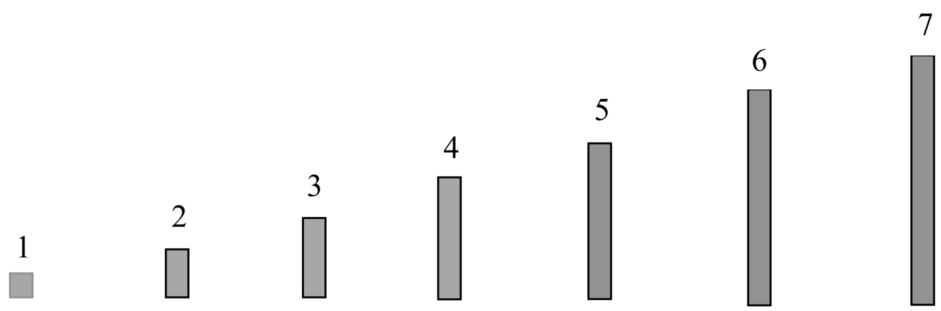
INSTRUCTIONS

You have been presented a list of word pairs on the computer screen. Now, you are asked to report the target words you remember from the list. The target words are the "*WORDS ON THE* ***RIGHT*** *SIDE IN THE PAIRS.*" To remember the words, utilise the "*CUES*, the left hand side words in the pairs studied, given to you on the next pages.

Importantly, you are required to give a response to every single cue, even if you have to guess. Write down your responses in the "**Targets**" column. Sometimes you may prefer to "**pass**" because you can't remember the target that was presented with the cue. If so, still write down your best guess, but check the "**pass**" checkbox next to that item. For all other responses, check the "**report**" checkbox, indicating that you are comfortable giving an answer to that item.

IMPORTANT: A response will be counted as correct even if it is a target paired with the wrong cue. It does not matter whether the target you write down is paired with the same cue that it was presented with at study. In other words, don't worry about matching the targets with the right cues.

In addition to writing down a response to every single item and checking the appropriate checkbox, also rate your confidence using the following 1-7 scale:



Not at all confident correct | Fairly confident correct | Completely confident correct

Reporting sheets for liberated cued-recall group to be used for the 2nd list

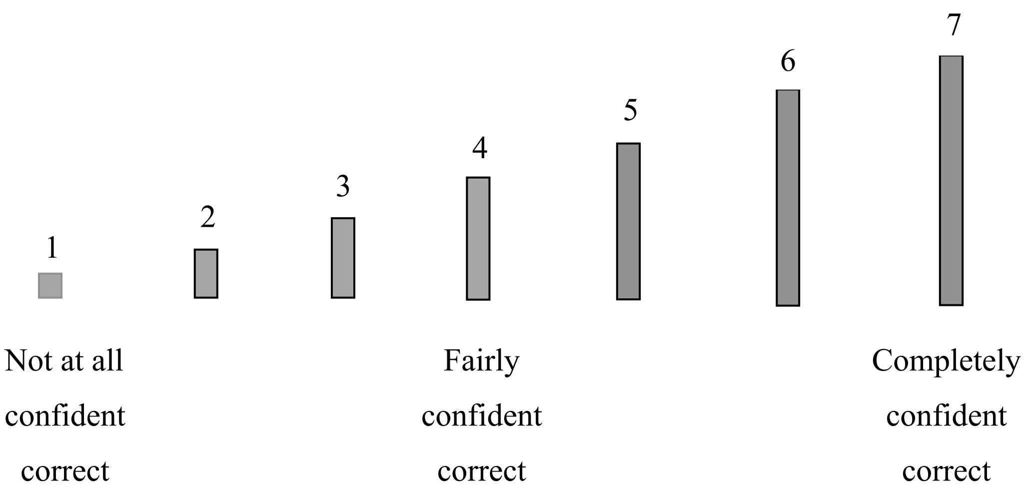
INSTRUCTIONS

You have been presented a second list of word pairs on the computer screen. Now, you are also asked to report the target words you remember from the list. The target words are the "*WORDS ON THE* ***RIGHT*** *SIDE IN THE PAIRS.*" To remember the words from the second list, again use the "*CUES*" given to you on the next pages.

Importantly, you are again required to give a response to every single cue, even if you have to guess. Write down your responses in the "**Targets**" column. Sometimes you may prefer to "**pass**" because you can't remember the target that was presented with the cue. If so, still write down your best guess, but check the "**pass**" checkbox next to that item. For all other responses, check the "**report**" checkbox, indicating that you are comfortable giving an answer to that item.

IMPORTANT: Again, a response will be counted as correct even if it is a target paired with the wrong cue. It does not matter whether the target you write down is paired with the same cue that it was presented with at study. In other words, don't worry about matching the targets with the right cues.

In addition to writing down a response to every single item and checking the appropriate checkbox, also rate your confidence using the following 1-7 scale:



CUES	TARGETS	REPORT	PASS	CONFIDENCE						
SCHOOL	____________	*Report* ❑	*Pass* ❑	1	2	3	4	5	6	7
VICTIM	____________	*Report* ❑	*Pass* ❑	1	2	3	4	5	6	7
TENNIS	____________	*Report* ❑	*Pass* ❑	1	2	3	4	5	6	7
MOVIE	____________	*Report* ❑	*Pass* ❑	1	2	3	4	5	6	7
STREET	____________	*Report* ❑	*Pass* ❑	1	2	3	4	5	6	7
SHOWER	____________	*Report* ❑	*Pass* ❑	1	2	3	4	5	6	7
PEOPLE	____________	*Report* ❑	*Pass* ❑	1	2	3	4	5	6	7
TOWEL	____________	*Report* ❑	*Pass* ❑	1	2	3	4	5	6	7
DRUM	____________	*Report* ❑	*Pass* ❑	1	2	3	4	5	6	7
CALENDAR	____________	*Report* ❑	*Pass* ❑	1	2	3	4	5	6	7
PLANE	____________	*Report* ❑	*Pass* ❑	1	2	3	4	5	6	7
GARBAGE	____________	*Report* ❑	*Pass* ❑	1	2	3	4	5	6	7
DOCTOR	____________	*Report* ❑	*Pass* ❑	1	2	3	4	5	6	7
CAR	____________	*Report* ❑	*Pass* ❑	1	2	3	4	5	6	7
MUSIC	____________	*Report* ❑	*Pass* ❑	1	2	3	4	5	6	7
BOTTLE	____________	*Report* ❑	*Pass* ❑	1	2	3	4	5	6	7
STAR	____________	*Report* ❑	*Pass* ❑	1	2	3	4	5	6	7
MUSCLE	____________	*Report* ❑	*Pass* ❑	1	2	3	4	5	6	7
DOLL	____________	*Report* ❑	*Pass* ❑	1	2	3	4	5	6	7
CAKE	____________	*Report* ❑	*Pass* ❑	1	2	3	4	5	6	7

CUES	TARGETS	REPORT	PASS	CONFIDENCE
CIRCUS	______________	*Report* ❑	*Pass* ❑	1 2 3 4 5 6 7
OCEAN	______________	*Report* ❑	*Pass* ❑	1 2 3 4 5 6 7
PUPPET	______________	*Report* ❑	*Pass* ❑	1 2 3 4 5 6 7
HONEY	______________	*Report* ❑	*Pass* ❑	1 2 3 4 5 6 7
VASE	______________	*Report* ❑	*Pass* ❑	1 2 3 4 5 6 7
MALL	______________	*Report* ❑	*Pass* ❑	1 2 3 4 5 6 7
SKIN	______________	*Report* ❑	*Pass* ❑	1 2 3 4 5 6 7
LAKE	______________	*Report* ❑	*Pass* ❑	1 2 3 4 5 6 7
PERFUME	______________	*Report* ❑	*Pass* ❑	1 2 3 4 5 6 7
AGENDA	______________	*Report* ❑	*Pass* ❑	1 2 3 4 5 6 7
ISLAND	______________	*Report* ❑	*Pass* ❑	1 2 3 4 5 6 7
RAZOR	______________	*Report* ❑	*Pass* ❑	1 2 3 4 5 6 7
NAIL	______________	*Report* ❑	*Pass* ❑	1 2 3 4 5 6 7
BIKE	______________	*Report* ❑	*Pass* ❑	1 2 3 4 5 6 7
MASK	______________	*Report* ❑	*Pass* ❑	1 2 3 4 5 6 7
DICE	______________	*Report* ❑	*Pass* ❑	1 2 3 4 5 6 7
BABY	______________	*Report* ❑	*Pass* ❑	1 2 3 4 5 6 7
CHESS	______________	*Report* ❑	*Pass* ❑	1 2 3 4 5 6 7
TOBACCO	______________	*Report* ❑	*Pass* ❑	1 2 3 4 5 6 7
LENS	______________	*Report* ❑	*Pass* ❑	1 2 3 4 5 6 7

Reporting sheets for the uncued-recall group to be used for the 1st list

Before you read the instructions, please fill the sections below:

Age: _______Gender: M____F ____

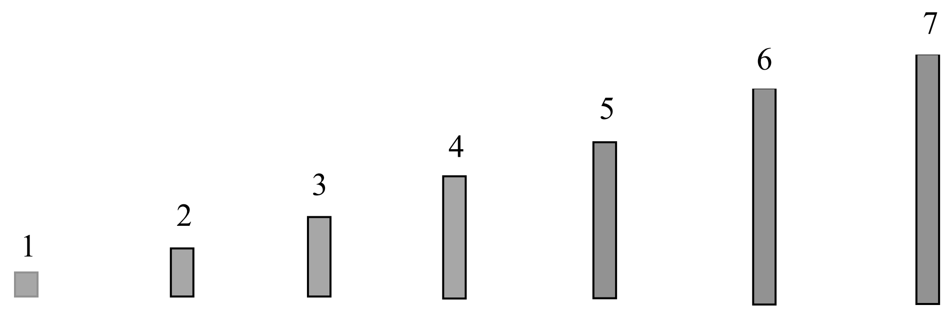
School: ______________________

INSTRUCTIONS

You have been presented a list of word pairs on the computer screen. Now, you are asked to report the target words you remember from the list. The target words are the "*WORDS ON THE* ***RIGHT*** *SIDE IN THE PAIRS.*" **Therefore, please try to remember and report only those words which you think are target words**.

Importantly, you are required to give a response, even if you have to guess. You are asked to give totally 20 words which you think are targets. Write down your responses in the "**Targets**" column. Sometimes you may prefer to "**pass**" because you can't remember the target. If so, still write down your best guess, but check the "**pass**" checkbox next to that item. For all other responses, check the "**report**" checkbox, indicating that you are comfortable giving an answer to that item.

In addition to writing down a response to every single item and checking the appropriate checkbox, also rate your confidence using the following 1-7 scale:



Not at all confident correct	Fairly confident correct	Completely confident correct

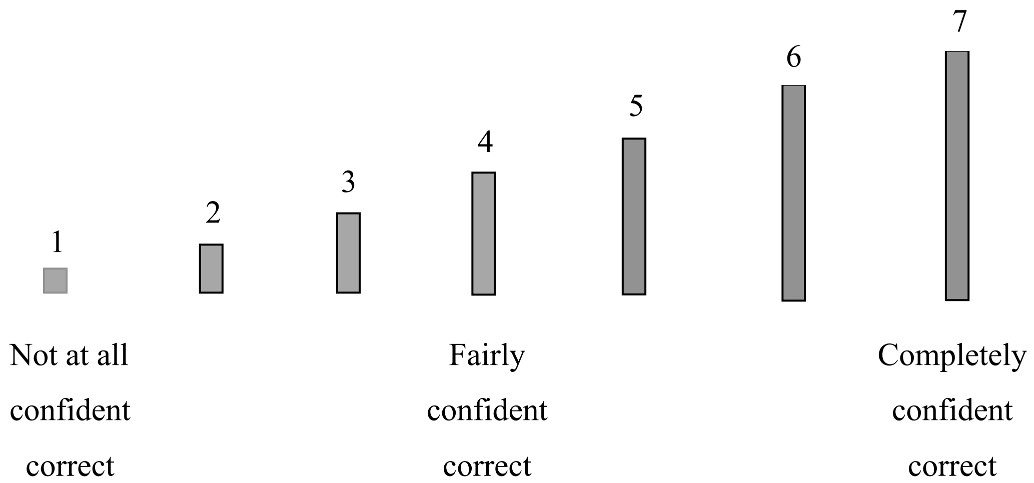
Reporting sheets for the uncued-recall group to be used for the 2nd list

INSTRUCTIONS

You have been presented a second list of word pairs on the computer screen. Now, you are also asked to report the target words you remember from the list. The target words are the "*WORDS ON THE* ***RIGHT*** *SIDE IN THE PAIR*". **Hence, please try to remember and report only words those from the second list that you think are target words.**

Importantly, you are again required to give a response, even if you have to guess. You are again asked to give totally 20 words which you think are targets. Write down your responses in the "**Targets**" column. Sometimes you may prefer to "**pass**" because you can't remember the target. If so, still write down your best guess, but check the "**pass**" checkbox next to that item. For all other responses, check the "**report**" checkbox, indicating that you are comfortable giving an answer to that item.

In addition to writing down a response to every single item and checking the appropriate checkbox, also rate your confidence using the following 1-7 scale:



	TARGETS	REPORT	PASS	CONFIDENCE
1)	____________	*Report* ❑	*Pass* ❑	1 2 3 4 5 6 7
2)	____________	*Report* ❑	*Pass* ❑	1 2 3 4 5 6 7
3)	____________	*Report* ❑	*Pass* ❑	1 2 3 4 5 6 7
4)	____________	*Report* ❑	*Pass* ❑	1 2 3 4 5 6 7
5)	____________	*Report* ❑	*Pass* ❑	1 2 3 4 5 6 7
6)	____________	*Report* ❑	*Pass* ❑	1 2 3 4 5 6 7
7)	____________	*Report* ❑	*Pass* ❑	1 2 3 4 5 6 7
8)	____________	*Report* ❑	*Pass* ❑	1 2 3 4 5 6 7
9)	____________	*Report* ❑	*Pass* ❑	1 2 3 4 5 6 7
10)	____________	*Report* ❑	*Pass* ❑	1 2 3 4 5 6 7
11)	____________	*Report* ❑	*Pass* ❑	1 2 3 4 5 6 7
12)	____________	*Report* ❑	*Pass* ❑	1 2 3 4 5 6 7
13)	____________	*Report* ❑	*Pass* ❑	1 2 3 4 5 6 7
14)	____________	*Report* ❑	*Pass* ❑	1 2 3 4 5 6 7
15)	____________	*Report* ❑	*Pass* ❑	1 2 3 4 5 6 7
16)	____________	*Report* ❑	*Pass* ❑	1 2 3 4 5 6 7
17)	____________	*Report* ❑	*Pass* ❑	1 2 3 4 5 6 7
18)	____________	*Report* ❑	*Pass* ❑	1 2 3 4 5 6 7
19)	____________	*Report* ❑	*Pass* ❑	1 2 3 4 5 6 7
20)	____________	*Report* ❑	*Pass* ❑	1 2 3 4 5 6 7

Debriefing Statement

The aim of the research was to investigate the effects of study type and report type as well as the inter-target association level on memory and metamemory performances. It is expected that target-focused way of study would have a dissociation between memory and metamemory performance as a function of high ITA level. The lowest quantity and the highest accuracy performance is expected for those subjects study pairs having low-ITA level and asked to recall targets without any cue help. Your data will help our understanding of how the memory performances are affected by the inter-target association level as a contextual effect at the time of study and by the effect of reporting option. The experiment did not use any deception. You may have a copy of this summary and if you wish you may also learn the summary of research findings with the following contact information.

If you have any further questions please contact me, Mehmet Akif Guzel, by the following e-mail address: **mag4v07@soton.ac.uk.**
Thank you for your participation in this research.

Signature _________________________________ Date ___________________
Name Mehmet Akif Guzel

I understand that if I have questions about my rights as a participant in this research, or if I feel that I have been placed at risk, I can contact the Chair of the Ethics Committee, Department of Psychology, University of Southampton, Southampton, SO17 1BJ. Phone: (023) 8059 5578.

Appendix D

Materials used in Experiment

Table 20

Twenty-four Categories of Exemplars that Constructed the Targets Pool (Totally 156 Words)

	Article of clothing	Cr.	Fr.	Nr.	A fish	Cr.	Fr.	Nr.
1	Shirt	6.05	27	90	Salmon	624	3	51
2	Pants	6.15	9	85	Trout	593	4	51
3	Socks	5.46	7	76	Goldfish	-	-	44
4	Shoe	5.96	14	53	Tuna	653	0	25
5	Hat	5.97	56	44	Catfish	-	-	27
6	Short	5.47	212	34	Shark	598	3	24
7	Jacket	6.31	33	33	Swordfish	-	-	15
8	Skirt	6.28	21	30	Herring	-	-	11
9	Jeans	6.35	1	20	Cod	-	-	9
10	Coat	5.83	43	19	Dolphin	0	1	8
11	Gloves	6.14	7	17	Piranha	-	-	5
12	Scarf	4.07	4	14	Whale	623	0	5
13	Blouse	6.36	1	12	Guppy	-	-	7
	A type of reading material				A musical instrument			
1	Magazine	-	-	93	Drum	-	-	75
2	Book	6.09	193	92	Guitar	6.73	19	72
3	Newspaper	6.56	65	71	Flute	5.66	1	71
4	Novel	5.34	59	26	Piano	6.26	38	60
5	Journal	5.50	42	21	Trumpet	7.00	7	55
6	Article	5.82	68	17	Clarinet	5.66	1	46
7	Textbook	-	-	17	Violin	6.13	11	44
8	Pamphlet	-	-	13	Trombone	6.02	0	35
9	Comics	-	-	6	Tuba	0.00	1	33
10	Encyclopaedia	0	16	6	Cello	0.00	0	21
11	Essay	5.22	19	6	Oboe	0.00	0	17
12	Flyer	-	-	6	Harp	6.94	1	10
13	Letter	5.16	145	6	Harmonica	-	-	6
	A male first name				A type of car			
1	John	-	-	45	Ford	-	-	44
2	Mike	-	-	36	Honda	-	-	41
3	Bob	-	-	26	Toyota	-	-	31
4	Matt	-	-	26	Chevrolet	-	-	27
5	Chris	-	-	25	Mercedes	-	-	25
6	Joe	-	-	19	Jeep	-	-	19
7	Brian	-	-	18	Porsche	-	-	15
8	Tom	-	-	18	Lexus	-	-	14
9	Steve	-	-	16	Nissan	-	-	14
10	Dan	-	-	15	Dodge	-	-	13
11	David	-	-	15	Mustang	-	-	13
12	Mark	-	-	15	Ferrari	-	-	11
13	James	-	-	13	Audi	-	-	9

Table 20 (*continued*).

	A relative	Cr.	Fr.	Nr.	A flower	Cr.	Fr.	Nr.
1	Uncle	5.71	57	95	Rose	5.86	86	96
2	Aunt	5.79	22	93	Daisy	6.09	3	59
3	Cousin	0.00	51	85	Tulip	6.05	4	53
4	Mother	5.47	216	80	Lily	6.53	1	30
5	Father	5.90	183	77	Carnation	-	-	29
6	Grandma	-	-	41	Daffodil	-	-	23
7	Grandpa	-	-	36	Dandelion	5.38	1	20
8	Sister	5.71	38	74	Sunflower	-	-	23
9	Brother	5.91	78	72	Pansy	-	-	11
10	Niece	0.00	8	32	Orchid	5.86	3	9
11	Nephew	6.19	9	30	Petunia	-	-	9
12	Son	6.38	166	8	Lilac	-	-	8
13	Daughter	6.25	72	6	Columbine	-	-	5
	A sport				A metal			
1	Football	5.75	36	87	Steel	0.00	9	62
2	Basketball	5.88	9	75	Iron	5.47	43	57
3	Soccer	5.05	3	75	Silver	5.36	29	55
4	Baseball	5.75	57	73	Copper	4.98	13	53
5	Tennis	5.54	15	54	Gold	5.68	52	53
6	Hockey	5.23	1	45	Aluminium	5.19	18	43
7	Golf	6.10	34	29	Platinum	-	-	17
8	Volleyball	6.48	1	29	Tin	5.87	12	15
9	Softball	0.00	0	19	Bronze	6.57	11	14
10	Rugby	-	-	19	Nickel	5.84	7	13
11	Polo	-	-	7	Lead	4.98	129	10
12	Bowling	0.00	0	6	Brass	5.53	19	9
13	Cricket	0.00	3	5	Zinc	-	-	9
	A colour				A flavouring substance			
1	Blue	4.49	143	100	Salt	5.69	46	87
2	Red	4.97	197	96	Pepper	5.61	13	85
3	Green	5.46	116	93	Garlic	6.23	4	29
4	Yellow	5.18	55	92	Oregano	5.39	0	19
5	Purple	4.27	13	83	Cinnamon	5.95	0	17
6	Black	4.66	203	77	Paprika	-	-	14
7	White	4.68	365	60	Basil	-	-	11
8	Pink	4.08	48	54	Vanilla	5.38	1	11
9	Brown	4.63	176	40	Mustard	5.64	20	10
10	Gray	5.06	12	26	Thyme	-	-	6
11	Violet	5.12	7	20	Curry	-	-	5
12	Maroon	-	-	8	Nutmeg	-	-	5
13	Indigo	-	-	11	Parsley	0.00	1	5

Table 20 (*continued*).

	A natural earth formation	Cr.	Fr.	Nr.	A fruit	Cr.	Fr.	Nr.
1	Mountain	6.25	33	83	Apple	7.00	9	95
2	River	5.83	16	39	Banana	6.29	4	71
3	Ocean	5.63	34	35	Grape	5.85	3	52
4	Volcano	6.83	2	35	Pear	6.30	6	50
5	Lake	5.70	54	34	Peach	6.05	3	40
6	Valley	6.66	73	30	Strawberry	7.00	2	40
7	Hill	6.02	72	26	Kiwi	-	-	30
8	Rock	6.03	75	23	Pineapple	-	-	26
9	Canyon	0	12	20	Watermelon	-	-	24
10	Plateau	-	-	15	Plum	6.18	2	21
11	Cave	5.79	9	8	Mango	-	-	18
12	Island	6.40	167	8	Cherry	5.86	6	15
13	Cliff	5.98	11	7	Apricot	-	-	5
	An article of furniture				A vegetable			
1	Chair	6.12	66	90	Carrot	6.09	1	77
2	Table	6.00	198	75	Lettuce	5.67	0	49
3	Couch	5.74	12	70	Broccoli	0.00	1	42
4	Bed	5.15	127	58	Tomato	6.85	4	36
5	Desk	5.79	65	49	Cucumber	6.49	0	31
6	Sofa	6.25	6	32	Peas	-	-	31
7	Lamp	6.24	7	17	Potato	7.00	15	28
8	Dresser	5.56	1	28	Celery	6.30	4	27
9	Stool	-	-	11	Onion	6.16	15	24
10	Futon	-	-	8	Spinach	6.90	2	14
11	Armoire	-	-	7	Bean	-	-	12
12	Cabinet	6.18	17	7	Cabbage	6.07	4	10
13	Bookshelf	-	-	6	Radish	0.00	8	10
	A transportation vehicle				A kitchen utensil			
1	Car	6.35	274	89	Knife	6.08	76	95
2	Bus	6.53	34	58	Fork	5.25	14	93
3	Truck	7.00	57	56	Spoon	5.88	6	93
4	Plane	5.31	114	54	Spatula	5.74	0	55
5	Train	5.79	82	44	Pan	5.74	16	22
6	Bicycle	6.33	5	42	Pot	5.30	33	20
7	Van	6.20	32	32	Blender	-	-	14
8	Boat	6.33	72	30	Bowl	5.26	23	14
9	Taxi	6.28	16	10	Plate	5.74	22	13
10	Subway	0.00	7	10	Ladle	-	-	14
11	Motorcycle	0.00	0	30	Tongs	-	-	8
12	Helicopter	-	-	8	Stove	5.75	15	5
13	Ship	6.25	83	30	Mixer	-	-	7

Table 20 (*continued*).

	A natural earth formation	Cr.	Fr.	Nr.	A fruit	Cr.	Fr.	Nr.
1	Dog	5.75	15	98	Cotton	6.28	38	96
2	Cat	6.21	23	97	Silk	5.26	12	70
3	Horse	6.03	117	52	Wool	6.02	10	50
4	Lion	6.17	17	41	Nylon	6.12	1	26
5	Bear	5.81	57	37	Satin	5.90	5	17
6	Tiger	6.07	7	36	Spandex	-	-	17
7	Cow	6.12	29	35	Denim	0.00	0	14
8	Elephant	7.00	7	28	Leather	5.71	24	13
9	Deer	-	-	23	Lycra	-	-	5
10	Mouse	6.12	10	23	Velvet	5.68	4	10
11	Pig	6.92	8	21	Suede	5.74	0	8
12	Giraffe	0.00	0	16	Cashmere	-	-	7
13	Rabbit	6.04	11	14	Flannel	5.46	4	5
	An occupation				An alcoholic beverage			
1	Doctor	5.75	100	72	Beer	5.83	34	87
2	Teacher	6.38	80	66	Vodka	-	-	62
3	Lawyer	5.53	43	54	Wine	6.40	72	54
4	Nurse	5.48	17	25	Rum	5.96	3	43
5	Fireman	-	-	14	Whiskey	6.00	17	32
6	Professor	6.52	57	14	Tequila	-	-	24
7	Dentist	6.03	12	12	Liquor	6.26	43	11
8	Engineer	0.00	42	10	Gin	6.35	23	23
9	Manager	6.04	88	9	Bacardi	-	-	6
10	Policeman	6.69	155	6	Champagne	6.12	13	9
11	Secretary	5.58	191	10	Martini	-	-	6
12	Cook	4.91	47	7	Smirnoff	-	-	6
13	Carpenter	0.00	6	5	Margarita	-	-	12
	A part of human body				A country			
1	Leg	6.04	58	87	America	0.00	194	90
2	Arm	5.53	94	82	Canada	-	-	56
3	Foot	3.46	70	71	France	-	-	53
4	Finger	6.20	40	67	England	-	-	38
5	Head	5.98	424	61	Mexico	-	-	52
6	Toe	5.96	9	61	Germany	-	-	38
7	Eye	6.28	122	59	Spain	-	-	33
8	Hand	5.60	431	54	Italy	-	-	29
9	Nose	4.98	60	53	China	5.74	69	26
10	Ear	6.26	29	49	Japan	-	-	23
11	Mouth	5.47	30	38	Russia	-	-	23
12	Stomach	6.04	37	35	Ireland	-	-	16
13	Heart	6.02	173	27	Greece	-	-	6

Note. Cr. = concreteness value; Fr. = written frequency; Nr. = Free association norms which display the percentage (%) of 642 participants who gave the response as an exemplar of the particular category asked in the norms study of Van Overschelde, Rawson, and Dunlosky (2004). Written frequency (Francis & Kucera) and concreteness values of the words based on the MRC Psycholinguistic Database: Machine Usable Dictionary (*source:* http://www.psych.rl.ac.uk/MRC_Psych_Db_files/mrc2.html).

Table 21

The Words Constituted the Cues Pool

	Cr.	Fr.		Cr.	Fr.		Cr.	Fr.
School	5.25	492	Cabin	6.23	23	Chess	0.00	3
Victim	5.49	27	Shelter	5.60	70	Puppet	-	-
Illusion	2.03	97	Customer	-	-	Guilt	2.92	33
Movie	5.85	29	Planet	5.64	21	History	3.03	286
Street	5.84	244	Citizen	-	-	Risk	2.70	54
Shower	5.88	15	Symbol	3.95	54	Error	2.85	80
People	5.51	847	Harbour	0.00	37	Blind	4.39	47
Towel	6.83	6	Urban	-	-	Empty	3.65	64
Impact	3.32	67	Avenue	6.48	46	Motion	3.63	55
Calendar	6.48	28	Cream	6.08	20	Report	5.67	174
Mask	6.38	9	Degree	-	-	Guest	5.15	39
Garbage	5.68	7	Prize	5.26	28	Valley	6.66	73
Empire	-	-	Sand	6.25	28	Name	3.96	294
Box	5.91	70	Concrete	5.50	48	Test	5.08	119
Effort	2.22	145	Future	2.35	227	Factory	6.87	32
Bottle	6.94	76	Youth	4.12	82	Note	4.36	127
Star	5.59	25	Lesson	4.19	29	Oxygen	5.43	43
Muscle	6.00	42	Jungle	6.28	20	Husband	5.45	131
Doll	5.36	10	Crisis	2.81	82	Shadow	4.94	36
Cake	6.11	13	Flesh	6.90	52	Fashion	3.75	69
Bomb	3.34	36	Studio	-	-	Male	5.48	37
Grave	5.27	33	Ancient	3.12	69	Member	5.60	137
Signal	5.50	63	Cloud	5.42	28	Crowd	5.34	53
Moon	5.68	60	Retired	-	-	System	2.36	416

Note. Cr. = concreteness value; Fr. = written frequency; Nr. = Free association norms which display the percentage (%) of 642 participants who gave the response as an exemplar of the particular category asked in the norms study of Van Overschelde, Rawson, and Dunlosky (2004). Written frequency (Francis & Kucera) and concreteness values of the words based on the MRC Psycholinguistic Database: Machine Usable Dictionary (*source:* http://www.psych.rl.ac.uk/MRC_Psych_Db_files/mrc2.html).

Consent Form

I, Mehmet Akif Guzel, am a third year PhD student in the School of Psychology of the University of Southampton. I am requesting your participation in a study on memory performance. This will involve studying some separate lists of word pairs and then to remember the words from the just studied lists. The study will last between 40 to 50 minutes. Personal information will not be released to or viewed by anyone other than researchers involved in this project. Results of the study will not include your name or any other identifying characteristics.

Your participation is voluntary and you may withdraw your participation at any time. If you choose not to participate there will be no consequence to your grade or to your treatment as a student in the school. If you have any questions please ask them now or contact me, Mehmet Akif Guzel, via email (mag4v07@soton.ac.uk).

Signature: .... Date: ...... /...... / ....

Mehmet Akif Güzel

I... have read the above informed consent form.

I understand that I may withdraw my consent and discontinue participation at any time without penalty or loss of benefit to myself. I understand that data collected as part of this project will be treated confidentially, and that published results of this research project will maintain my confidentiality. In signing this consent letter, I am not waiving my legal claims, rights, or remedies. A copy of this letter will be offered to me.

Circle Yes or No

I give consent to participate in the above mentioned study. Yes No

Signature.. Date:....../....../..........

I understand that if I have questions about my rights as a participation in this research, or if I feel that I have been placed at risk, I can contact the Chair of the Ethics Committee, Department of Psychology, University of Southampton, Southampton, SO17 1BJ, Phone: (023) 8059 5578.

a) Instructions page used for Experiment 4 regardless of the group

INSTRUCTIONS

This is a study on memory. Throughout the experiment, you will be presented with 3 different lists of word pairs. After each of the presentation, you will be tested on remembering the TARGET words from the just presented list.

Each of the 3 lists contains word pairs. Each of the word pairs contains a CUE on the left hand side, and a TARGET on the right hand side.You will only be responsible for remembering the target words at test, but study the cue words as well because they will help you remember the targets later on."

Continue

b) Instructions page for the cued-recall group

At testing, there will be the columns of; CUES, TARGETS, ORDER, REPORT / PASS and CONFIDENCE LEVEL.

CUES = Words that were presenteed on the left hand side in the study list.

TARGETS = You need to write down the targets next to each cue. However, you should match the TARGET words with the SAME CUES paired at the time of presentation. Your response will be counted as correct ONLY WHEN you make the correct matchings.

ORDER = You can start reporting the target words in any order. That is, you can start from whicever cue word you wish to match the target with this particular cue word. However, write down the order of you reporting (starting from 1 and terminating at 24) under the ORDER column.

REPORT / PASS = You need to give a response to every single cue, even if you have to guess. Sometimes, you may prefer to PASS because you can not remember the target that was presented with the cues. If so, still write down your best guess, but check `PASS` checkbox next to that item. For all other responses, check the `REPORT` checkbox, indicating that you are comfortable giving an answer to that item.

CONFIDENCE LEVEL = In addition to writing down a response to every single item, indicating your order of reporting and then checking the appropriate check box, also rate your CONFIDENCE LEVEL indicating how confident you are that your response is correct. The scale will be from 1 to 7.

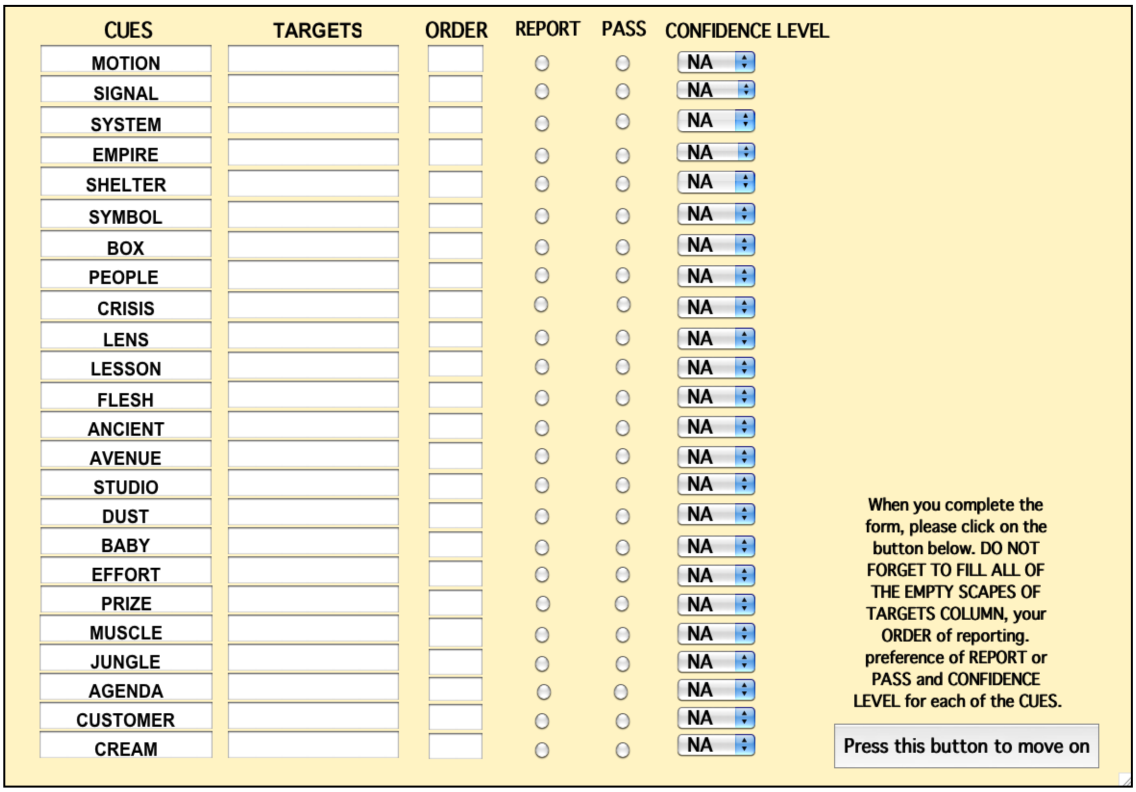
1 = not at all confident correct
2
3
4= fairly confident correct
5
6
7 = completely confident correct

Continue

Figure 17. Instructions page for the cued and the uncued-recall groups (a), instruction (b) and reporting pages (c) specific to the cued-recall group in Experiment 4.

Figure 17 (continued).

c) Reporting page for the cued-recall group

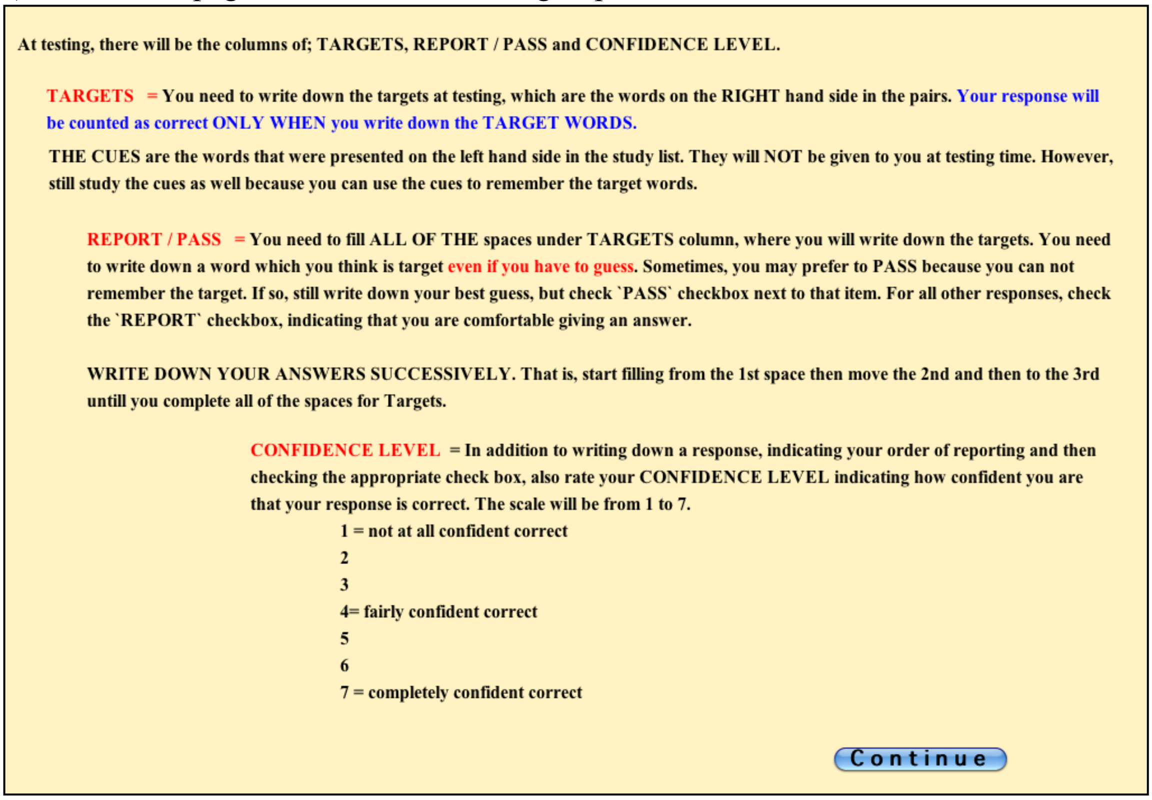


CUES	TARGETS	ORDER	REPORT	PASS	CONFIDENCE LEVEL
MOTION					NA
SIGNAL					NA
SYSTEM					NA
EMPIRE					NA
SHELTER					NA
SYMBOL					NA
BOX					NA
PEOPLE					NA
CRISIS					NA
LENS					NA
LESSON					NA
FLESH					NA
ANCIENT					NA
AVENUE					NA
STUDIO					NA
DUST					NA
BABY					NA
EFFORT					NA
PRIZE					NA
MUSCLE					NA
JUNGLE					NA
AGENDA					NA
CUSTOMER					NA
CREAM					NA

When you complete the form, please click on the button below. DO NOT FORGET TO FILL ALL OF THE EMPTY SCAPES OF TARGETS COLUMN, your ORDER of reporting. preference of REPORT or PASS and CONFIDENCE LEVEL for each of the CUES.

Press this button to move on

a) Instructions page for the uncued-recall group



At testing, there will be the columns of; TARGETS, REPORT / PASS and CONFIDENCE LEVEL.

TARGETS = You need to write down the targets at testing, which are the words on the RIGHT hand side in the pairs. Your response will be counted as correct ONLY WHEN you write down the TARGET WORDS.

THE CUES are the words that were presented on the left hand side in the study list. They will NOT be given to you at testing time. However, still study the cues as well because you can use the cues to remember the target words.

REPORT / PASS = You need to fill ALL OF THE spaces under TARGETS column, where you will write down the targets. You need to write down a word which you think is target even if you have to guess. Sometimes, you may prefer to PASS because you can not remember the target. If so, still write down your best guess, but check `PASS` checkbox next to that item. For all other responses, check the `REPORT` checkbox, indicating that you are comfortable giving an answer.

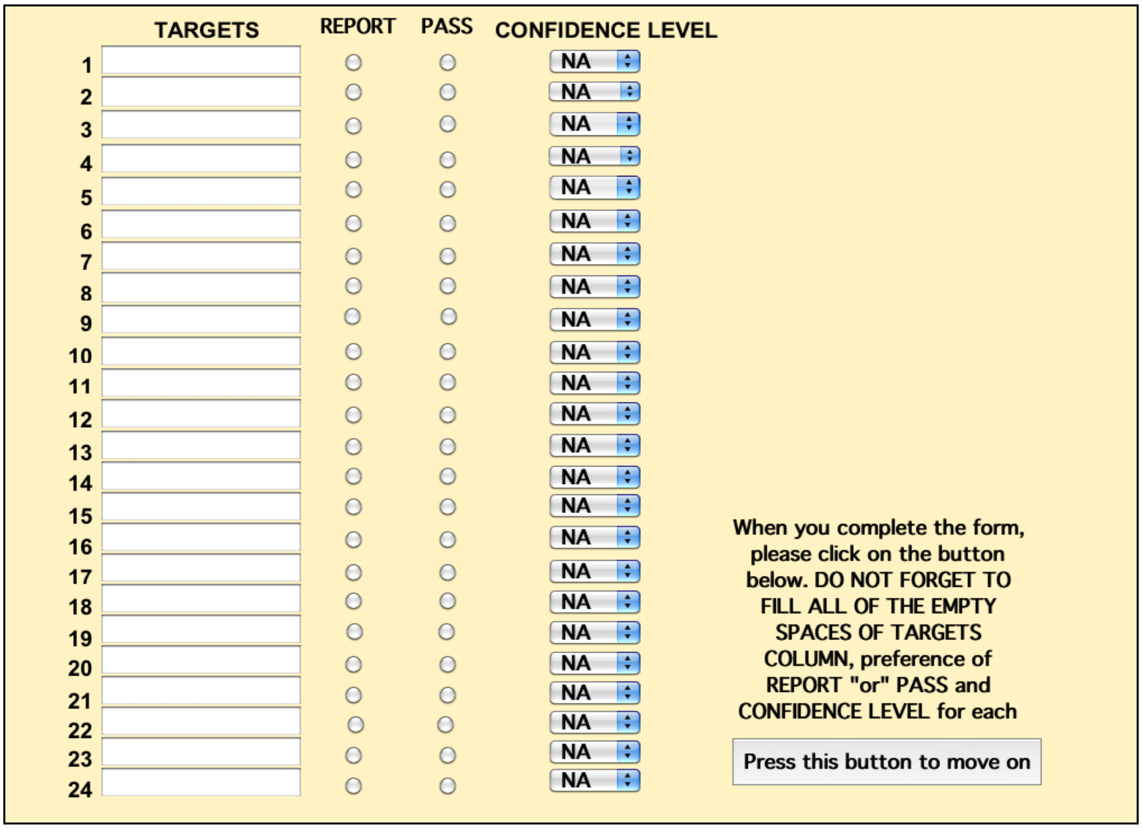
WRITE DOWN YOUR ANSWERS SUCCESSIVELY. That is, start filling from the 1st space then move the 2nd and then to the 3rd untill you complete all of the spaces for Targets.

CONFIDENCE LEVEL = In addition to writing down a response, indicating your order of reporting and then checking the appropriate check box, also rate your CONFIDENCE LEVEL indicating how confident you are that your response is correct. The scale will be from 1 to 7.

1 = not at all confident correct
2
3
4= fairly confident correct
5
6
7 = completely confident correct

Continue

b) Reporting page for the uncued-recall group



	TARGETS	REPORT	PASS	CONFIDENCE LEVEL
1		○	○	NA
2		○	○	NA
3		○	○	NA
4		○	○	NA
5		○	○	NA
6		○	○	NA
7		○	○	NA
8		○	○	NA
9		○	○	NA
10		○	○	NA
11		○	○	NA
12		○	○	NA
13		○	○	NA
14		○	○	NA
15		○	○	NA
16		○	○	NA
17		○	○	NA
18		○	○	NA
19		○	○	NA
20		○	○	NA
21		○	○	NA
22		○	○	NA
23		○	○	NA
24		○	○	NA

When you complete the form, please click on the button below. DO NOT FORGET TO FILL ALL OF THE EMPTY SPACES OF TARGETS COLUMN, preference of REPORT "or" PASS and CONFIDENCE LEVEL for each

Press this button to move on

Figure 18. Instruction (a) and reporting pages (b) for the uncued-recall group in Experiment 4.

Debriefing Statement

The aim of the research was to investigate the effects of inter-target association level on memory and metamemory performance. It was expected that participants would have higher memory performance when targets were inter-related to the highest extend, however, their monitoring performance would be lower for the inter-related targets compared to the unrelated ones. Your data will help our understanding of how the memory and metamemory performance are affected by the inter-target association level. The experiment did not use any deception. You may have a copy of this summary and if you wish you may also learn the summary of research findings with the following contact information.

If you have any further questions please contact me, Mehmet Akif Guzel, by the following e-mail address: **mag4v07@soton.ac.uk.**

Thank you for your participation in this research.

Signature ________________________________ Date ___________________

Name Mehmet Akif Guzel

I understand that if I have questions about my rights as a participant in this research, or if I feel that I have been placed at risk, I can contact the Chair of the Ethics Committee, Department of Psychology, University of Southampton, Southampton, SO17 1BJ. Phone: (023) 8059 5578.

Appendix E

Materials used in Experiment 5

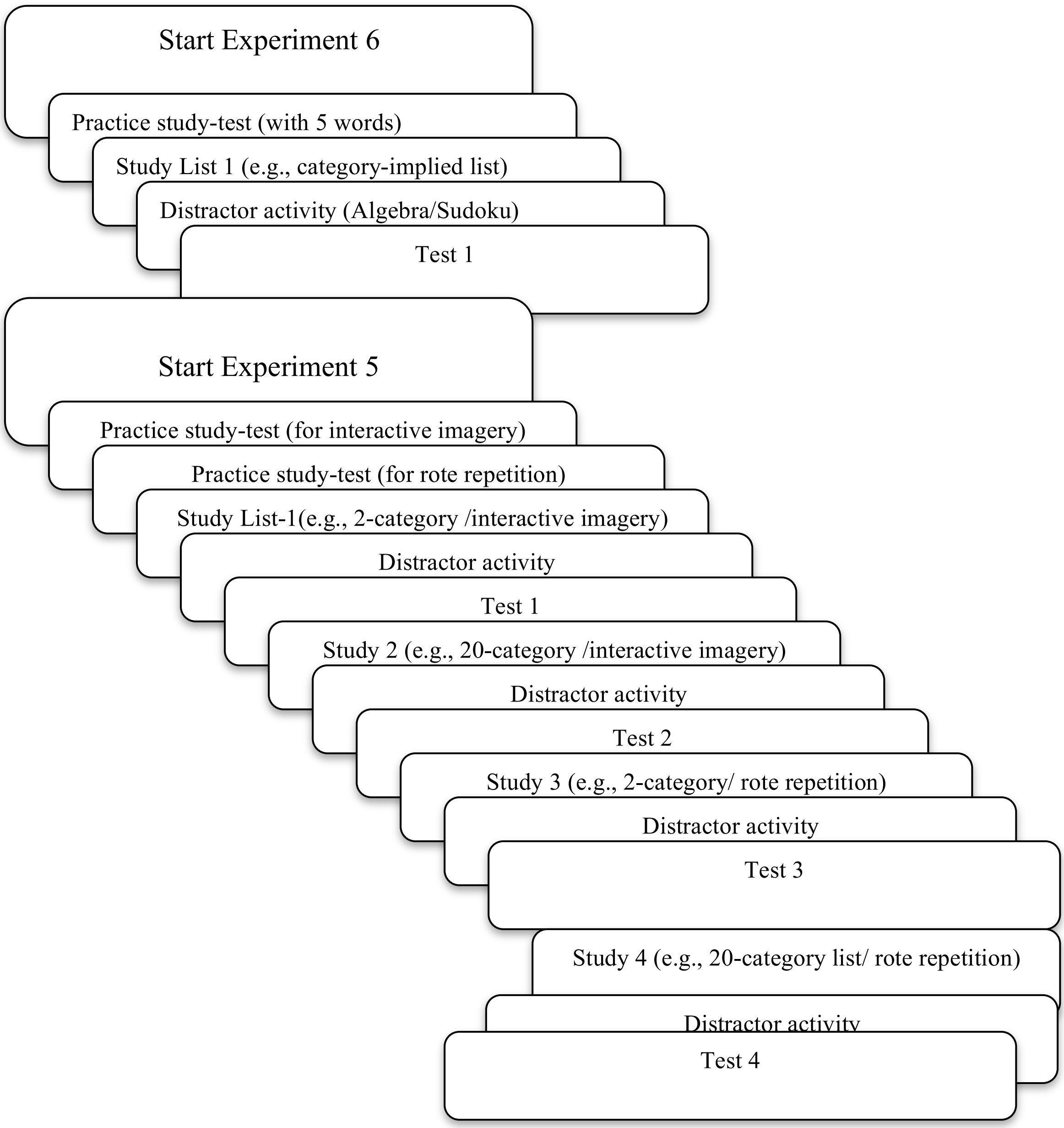


Figure 19. The Experimental procedure utilised so as to collect data for Experiments 5 and 6.

Table 22

Versions of study-test orders in Experiments 5 and 6.

Exp 6	Exp 5				Exp 6	Exp 5			
	Order of study/test cycles					Order of study/test cycles			
	1st	2nd	3rd	4th		1st	2nd	3rd	4th
CI	T1(i)	M1(i)	T2(r)	M2(r)	NCI	T1(i)	M1(i)	T2(r)	M2(r)
CI	T1(i)	M1(i)	M2(r)	T2(r)	NCI	T1(i)	M1(i)	M2(r)	T2(r)
CI	M1(i)	T1(i)	T2(r)	M2(r)	NCI	M1(i)	T1(i)	T2(r)	M2(r)
CI	M1(i)	T1(i)	M2(r)	T2 (r)	NCI	M1(i)	T1(i)	M2(r)	T2 (r)
CI	T1(r)	M1(r)	T2(i)	M2(i)	NCI	T1(r)	M1(r)	T2(i)	M2(i)
CI	T1(r)	M1(r)	M2(i)	T2(i)	NCI	T1(r)	M1(r)	M2(i)	T2(i)
CI	M1(r)	T1(r)	T2(i)	M2(i)	NCI	M1(r)	T1(r)	T2(i)	M2(i)
CI	M1(r)	T1(r)	M2(i)	T2(i)	NCI	M1(r)	T1(r)	M2(i)	T2(i)
CI	T2(i)	M2(i)	T1(r)	M1(r)	NCI	T2(i)	M2(i)	T1(r)	M1(r)
CI	T2(i)	M2(i)	M1(r)	T1(r)	NCI	T2(i)	M2(i)	M1(r)	T1(r)
CI	M2(i)	T2(i)	T1(r)	M1(r)	NCI	M2(i)	T2(i)	T1(r)	M1(r)
CI	M2(i)	T2(i)	M1(r)	T1(r)	NCI	M2(i)	T2(i)	M1(r)	T1(r)
CI	T2(i)	M2(i)	T1(r)	M1(r)	NCI	T2(i)	M2(i)	T1(r)	M1(r)
CI	T2(i)	M2(i)	M1(r)	T1(r)	NCI	T2(i)	M2(i)	M1(r)	T1(r)
CI	M2(i)	T2(i)	T1(r)	M1(r)	NCI	M2(i)	T2(i)	T1(r)	M1(r)
CI	M2(i)	T2(i)	M1(r)	T1(r)	NCI	M2(i)	T2(i)	M1(r)	T1(r)

Note. CI = category implied list; NCI = no category implied list; RR = rote repetition; IMG = interactive imagery; T = two-category study list; M = multiple-category study list; 1 = study list version 1; 2 = study list version 2. (i) = interactive imagery; (r) = rote repetition. The experimental versions displayed in the table is valid for cued recall group as well as for the uncued recall group. All of the participants in Experiment 5, regardless of the presentation order, started Experiment 5 after completing a single study-test cycle of Experiment 6. Hence, there was not a proactive or retroactive interference effect for Experiment 6, since the order of study-test cycles were counterbalanced for Experiment 5 and all participants completed Experiment 6. Therefore, a possible proactive interference effect of Experiment 6 on Experiment 5 was controlled for all of the participants.

Materials used in Experiment 5

Table 23

Study lists used in Experiment 5

	Cue	Fq.	Conc.	Img.	Target	Fr.A.	Fq.	Conc.	Img.
Two-category study list-1 (T1)									
1	Fireplace	6	592	639	Apple	95	9	620	637
2	Tulip	4	619	641	Banana	71	4	633	644
3	Drizzle	5	558	582	Cherry	15	6	611	582
4	Rectangle	4	554	590	Grape	52	3	611	591
5	Tooth	20	619	624	Orange	86	23	601	626
6	Menu	5	555	613	Peach	40	3	617	613
7	Witch	5	522	589	Pear	50	6	634	590
8	Feast	3	642	610	Plum	21	1	632	611
9	Zipper	1	599	632	Strawberry	40	2	610	631
10	Rocket	7	645	612	Tomato	23	4	662	610
	M	6	590.5	613.2		49.3	6.1	623.1	613.5
11	Uncle	57	580	574	Bear	37	57	585	572
12	Shower	15	588	615	Cat	97	23	615	617
13	Beard	26	580	630	Cow	35	29	621	632
14	Cake	13	624	624	Deer	23	13	631	624
15	Boat	72	637	631	Dog	98	75	610	630
16	Hammer	9	605	618	Elephant	28	7	628	616
17	Ball	110	615	622	Horse	52	117	613	624
18	Holiday	17	439	629	Lion	41	17	627	626
19	Microscope	8	591	617	Mouse	23	10	624	615
20	Coffin	7	595	606	Tiger	36	7	611	606
	M	33.4	585.4	616.6		47	35.5	616.5	616.2
Two-category Study list-2 (T2)									
1	Throat	51	578	561	Hat	44	56	601	562
2	Fence	30	597	611	Jacket	33	33	635	611
3	Brain	45	556	572	Coat	19	43	601	571
4	Mansion	8	579	628	Pants	85	9	619	630
5	Bullet	28	595	611	Shirt	90	27	616	612
6	Toilet	13	586	603	Shoe	53	14	600	601
7	Private	191	350	432	Short	34	212	351	431
8	Pupil	20	570	572	Skirt	30	21	614	573
9	Cruiser	4	571	553	Socks	76	4	581	553
10	China	69	597	597	Dress	19	67	595	595
	M	45.9	557.9	574		48.3	48.6	581.3	573.9
11	Hockey	1	535	593	Clarinet	46	1	633	593
12	Cliff	11	591	599	Drum	75	11	602	599
13	Measles	2	568	582	Flute	71	1	587	581
14	Helmet	1	602	620	Harp	10	1	591	621
15	Nursery	13	528	542	Bass	12	16	547	544
16	Quarter	34	509	531	Piano	60	38	615	630
17	Chalk	3	634	601	Saxophone	38	4	624	602
18	Bone	33	588	567	Horn	9	31	618	566
19	Mansion	8	579	628	Trumpet	55	7	608	628
20	Bubble	12	563	604	Violin	44	11	626	606
	M	11.8	569.7	586.7		42	12.1	605.1	597

Table 23 (*continued*).

	Cues	Fq	Conc	Img	Targets	Fr.A.	Fq	Conc	Img
	Multiple-category study list-1 (M1)								
1	Kiss	17	546	633	Mist	-	14	497	638
2	Sunlight	17	515	643	Cigarette	-	25	607	645
3	Trolley	5	590	585	Spoon	-	6	614	584
4	Noose	3	542	593	Spice	-	4	590	592
5	Autumn	22	421	622	Star	-	25	574	623
6	Pyramid	2	615	613	Gym	-	2	612	613
7	Brandy	7	595	590	Cradle	-	7	587	592
8	Mosquito	1	595	612	Scissors	-	1	596	609
9	Cork	9	608	631	Lobster	-	1	616	629
10	Broom	2	613	608	Hurricane	-	8	576	608
	M	8.5	564	613			9.3	586.9	613.3
11	Seat	54	568	574	Walking	-	54	497	574
12	Arrow	14	595	619	Photograph	-	18	590	618
13	Mountain	33	616	629	Christmas	-	27	432	629
14	Pillow	8	613	624	Diamond	-	8	610	623
15	Skin	47	614	638	Garden	-	60	602	635
16	Wallet	6	584	617	Corpse	-	7	587	614
17	Blood	121	613	620	Ball	-	110	615	622
18	Mirror	27	605	627	Policeman	-	19	574	629
19	Typewriter	10	611	615	Ankle	-	8	608	613
20	Coin	10	581	603	Saloon	-	12	575	608
	M	33	600	616.6			32.3	569	616.5
	Multiple-category study list-2 (M2)								
1	Concrete	48	562	564	Wire	-	42	585	564
2	Sister	38	575	613	Palace	-	38	579	612
3	Bay	57	580	570	Driver	-	49	553	567
4	Cafe	20	568	625	Sunburn	-	5	563	629
5	Honey	25	611	608	Chin	-	27	592	608
6	Infant	11	579	600	Fountain	-	18	593	602
7	Period	265	358	429	Period	-	283	379	432
8	Movie	29	590	571	Belly	-	23	630	576
9	Warrior	5	525	553	Bandage	-	4	639	554
10	Sheet	45	608	594	Engine	-	50	586	595
	M	54.3	555.6	572.7			53.9	569.9	573.9
11	Avalanche	1	554	596	Beggar	-	2	533	593
12	Rubber	15	596	599	Tower	-	13	585	596
13	Napkin	3	585	582	Dungeon	-	2	562	579
14	Kite	1	592	624	Jewel	-	1	594	621
15	Bow	15	572	546	Ambassador	-	22	546	545
16	Forest	66	609	633	Bedroom	-	50	607	628
17	Pendulum	2	583	605	Web	-	6	561	602
18	Banner	8	567	569	Aerial	-	8	517	567
19	Yacht	4	606	624	Volcano	-	2	591	627
20	Monk	16	570	606	Clover	-	16	554	606
	M	13.1	583.4	598.4			12.2	565	596.4

Note. Fq. = Francis-Kucera written frequency; Conc. = concreteness value; Img. = imageaility value; Fr.A. = free-association norms. Fq, Conc, and Img values were drawn from the MRC Psycholinguistic Database: Machine Usable Dictionary (*source:* http://www.psych.rl.ac.uk/MRC_Psych_Db_files/mrc2.html). Free association norms based on the norms study of Van Overschelde, Rawson, and Dunlosky (2004).

Consent Form

Information Sheet

I, Mehmet Akif Guzel, am a PhD student in the School of Psychology of the University of Southampton. I am requesting your participation in a study on memory. This will involve the presentations of several lists containing word pairs and the requirement to remember and report the words from the just presented list. The study will last between 50 and 60 minutes to complete. Personal information will not be released to or viewed by anyone other than researchers involved in this project. Results of the study will not include your name or any other identifying characteristics.

Your participation is voluntary and you may withdraw your participation at any time. If you choose not to participate there will be no consequence to your grade or to your treatment as a student in the psychology department. If you have any questions please ask them now, or contact me, Mehmet Akif Guzel, via the e-mail address: **mag4v07@soton.ac.uk.**

Signature Date

Name : Mehmet Akif Guzel

I_________________________________have read the above informed consent form. I understand that I may withdraw my consent and discontinue participation at any time without penalty or loss of benefit to myself. I understand data collected as part of this research project will be treated confidentially, and that published results of this research project will maintain my confidentiality. In signing this consent letter, I am not waiving my legal claims, rights, or remedies. A copy of this letter will be offered to me.

Circle Yes or No

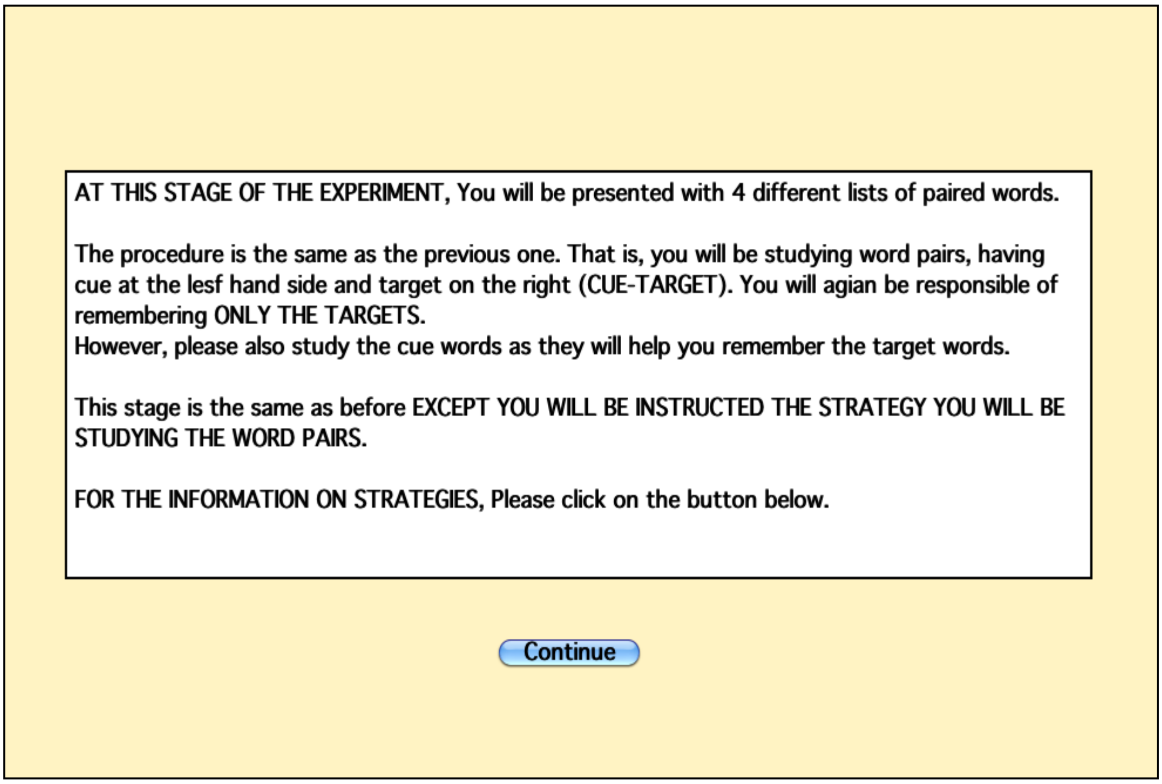
I give consent to participate in the above study. Yes No

Signature Date

Name

I understand that if I have questions about my rights as a participant in this research, or if I feel that I have been placed at risk, I can contact the Chair of the Ethics Committee, Department of Psychology, University of Southampton, Southampton, SO17 1BJ. Phone: (023) 8059 5578.

a) General instructions



b) Instructions page on how the encoding strategies would be implemented

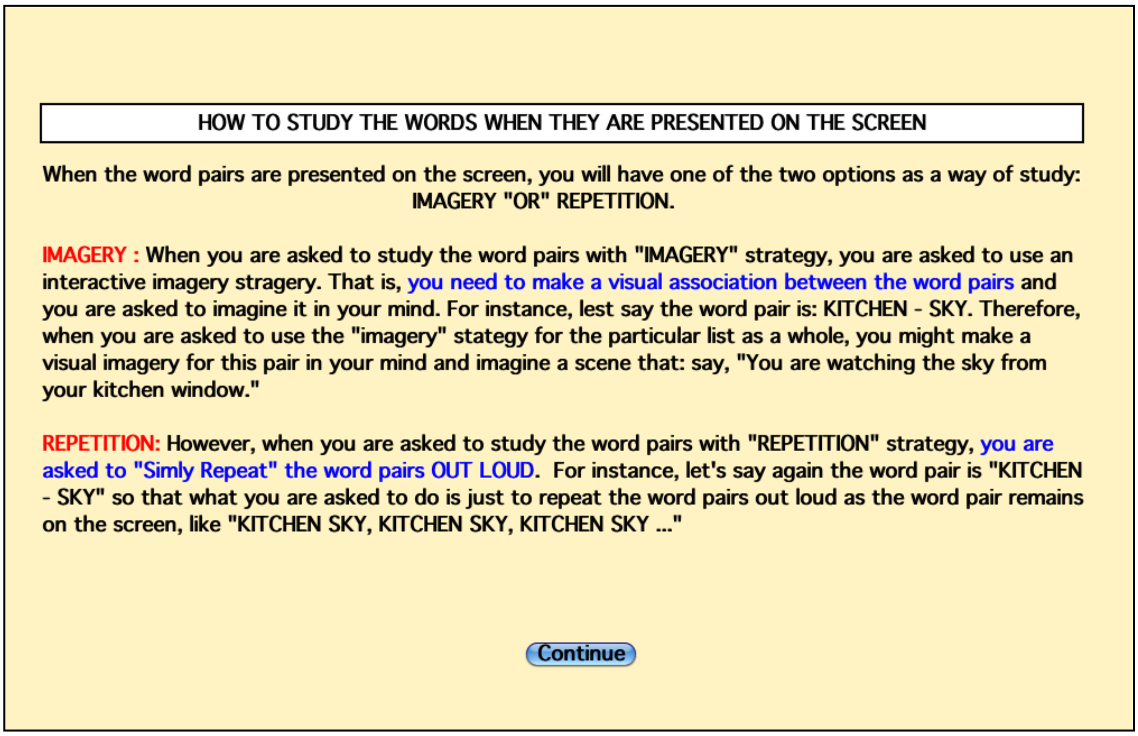
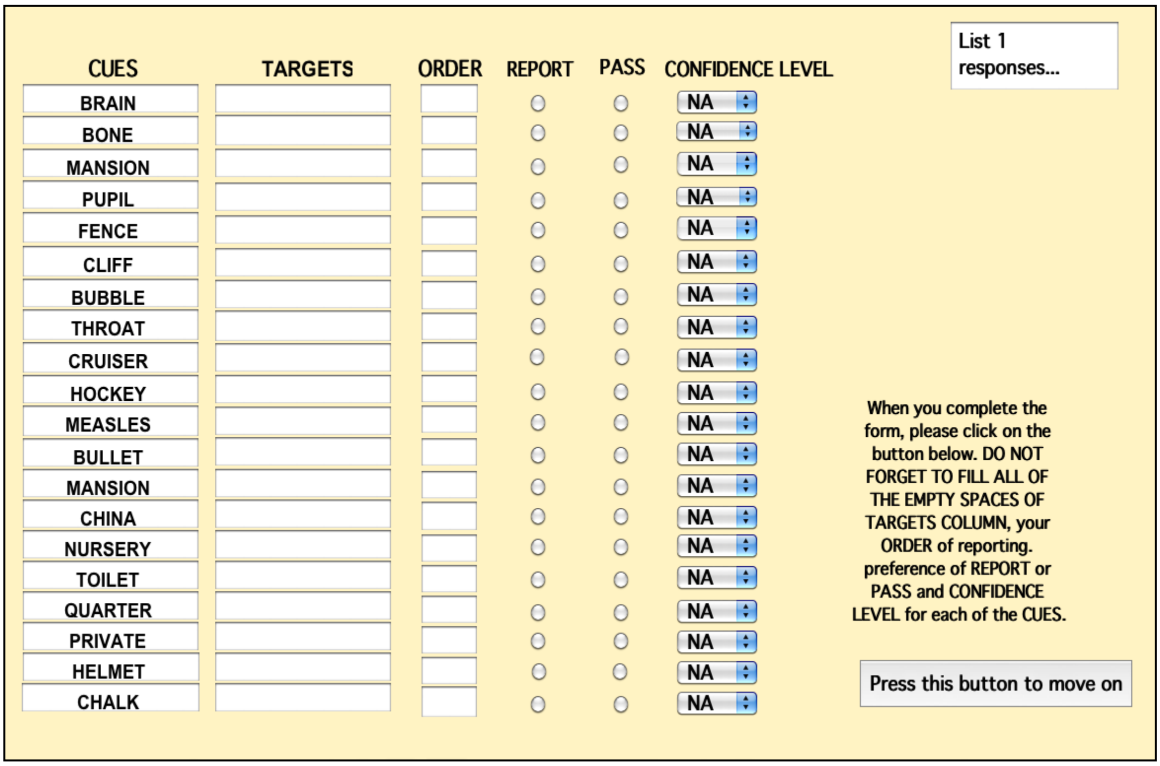


Figure 20. The pages of general instructions (a), instructions on how to implement the encoding strategies (b), and the reporting page for the cued-recall group (c) in Experiment 5.

Figure 20 (*continued*).

c) Reporting page for the cued-recall group



List 1 responses...

CUES	TARGETS	ORDER	REPORT	PASS	CONFIDENCE LEVEL
BRAIN			○	○	NA
BONE			○	○	NA
MANSION			○	○	NA
PUPIL			○	○	NA
FENCE			○	○	NA
CLIFF			○	○	NA
BUBBLE			○	○	NA
THROAT			○	○	NA
CRUISER			○	○	NA
HOCKEY			○	○	NA
MEASLES			○	○	NA
BULLET			○	○	NA
MANSION			○	○	NA
CHINA			○	○	NA
NURSERY			○	○	NA
TOILET			○	○	NA
QUARTER			○	○	NA
PRIVATE			○	○	NA
HELMET			○	○	NA
CHALK			○	○	NA

When you complete the form, please click on the button below. DO NOT FORGET TO FILL ALL OF THE EMPTY SPACES OF TARGETS COLUMN, your ORDER of reporting. preference of REPORT or PASS and CONFIDENCE LEVEL for each of the CUES.

Press this button to move on

Appendix F
Materials used in Experiment 6

Table 24

Category-implied study list and the no-category-implied study list used in Experiment 6

Category-implied study list						No-category-implied study list			
Cues	Targets	C.cn.	C.fr.	T.cn.	T.fr.	Cues	Targets	C.Cr.	C.Fr.
Hug	Arms	5.34	3	5.53	94	Weapons	Arms	6.38	42
Pain	Chest	4.22	88	5.76	53	Drawer	Chest	6.31	8
Listen	Ear	4.04	51	6.26	29	Corn	Ear	5.51	34
Sight	Eye	3.84	86	6.28	122	Needle	Eye	5.79	15
Wash	Hand	4.20	37	5.60	431	Help	Hand	2.82	311
Hat	Head	5.97	56	5.98	424	Tail	Head	6.09	24
Run	Leg	4.51	212	6.04	58	Table	Leg	6.00	198
Food	Mouth	5.84	147	5.47	30	River	Mouth	5.83	165
Broken	Neck	-	-	5.83	81	Bottle	Neck	6.94	76
Lick	Tongue	0.00	3	6.12	35	Language	Tongue	4.00	109
Dark	Black	4.68	185	4.66	203	Evil	Black	2.28	72
Hair	Brown	5.70	148	4.63	176	Gordon	Brown	-	-
Ocean	Blue	5.63	34	4.49	143	Sad	Blue	3.52	35
Dress	Pink	5.91	67	4.08	48	Elephant	Pink	7.00	7
Cloud	Gray	5.42	28	5.06	12	Old	Gray	3.61	660
Grass	Green	5.93	53	5.46	116	Amateur	Green	-	-
Sun	Orange	6.23	112	5.74	23	Tree	Orange	6.62	16
Rose	Red	5.86	86	4.97	197	Communist	Red	-	-
Leaf	Gold	5.89	12	5.68	52	Metal	Gold	6.76	61
Snow	White	6.05	59	4.68	365	British	White	-	-

Note. C.cn.= concreteness value of the cues; C.fr.= written frequency of the cues; T.cn. = concreteness value of the targets; T.fr.= written frequency of the targets; F.A.= forward semantic association (from target to cue); B.A.= backward semantic association (from cue to target). Written frequency (Francis & Kucera) and concreteness values of the words based on the MRC Psycholinguistic Database: Machine Usable Dictionary (*source:* http://www.psych.rl.ac.uk/MRC_Psych_Db_files/mrc2.html).

Piloting of Study Materials used in Experiment 6

Twenty-seven postgraduate students, who were all native English speakers, took part in the piloting of the study words in Experiment 6. Seven of them were male (25.9%) and 20 of them were female (74.1%; age: M = 26.2, SD = 1.9). In order to detect the target words, they filled out a form electronically, which asked them to write down first four words popped into their minds when they were the prompted with each of the 40 cue words. After this, they were again prompted with word the cue words and then asked to write down the first and the second word that come into their minds in order with a criterion to be followed: The freely-associated words had to be either word having a meaning of "a human body part" or "a type of colour". Table 25 displays the frequencies (%) of three most common responses given as a first or as a second response to the primed cues by following the category-restriction criterion at free association.

The form used to collect pilot date of study materials used in Experiment 6

Your age: _________ Gender: _________

For each of the words given below, please write down 4 related words that come to your mind; and please write them down in sequence of the words coming to your mind (1st word coming to your mind goes into column 1st, and 2nd word coming to your mind goes into column 2nd, and so on). Note: You *can* write down the same word(s) repeatedly for different prompted words. But please write down 4 different words for each of the prompted words.

For example:

Prompted word	1st	2nd	3rd	4th
PAPER	Magazine	Tree	Newspaper	Printer

THE WORDS THAT YOU ARE ASKED TO FREE ASSOCIATE ARE AS FOLLOW:

	Prompted word	1st	2nd	3rd	4th
1	LEAF				
2	SNOW				
3	WEAPONS				
4	WASH				
5	HAT				
6	RUN				
7	FOOD				
8	BROKEN				
9	LICK				
10	DARK				
11	OCEAN				
12	HAIR				
13	DRESS				
14	CLOUD				
15	GRASS				
16	SUN				
17	ROSE				
18	HUG				
19	PAIN				
20	LISTEN				

	Prompted word	1st	2nd	3rd	4th
1	**NEEDLE**				
2	**DRAWER**				
3	**CORN**				
4	**SIGHT**				
5	**HELP**				
6	**TAIL**				
7	**TABLE**				
8	**EVIL**				
9	**GORDON**				
10	**SAD**				
11	**LANGUAGE**				
12	**BRITISH**				
13	**TREE**				
14	**RIVER**				
15	**BOTTLE**				
16	**AMATEUR**				
17	**COMMUNIST**				
18	**METAL**				
19	**ELEPHANT**				
20	**OLD**				

Now, you are asked to write down first 2 words, which you think that they are associated to the words given (in the order that they come to your mind, 1^{s} & 2^{nd}). However, there is a <u>very important criterion that you are asked to follow: The words that you will write down must be related to the word given in terms of the following categories: either A HUMAN BODY PART or A TYPE OF COLOUR.</u>

(Note. You *can* write down the same words for different prompted words. Nevertheless, please think about the prompted words individually).

For example:

A HUMAN BODY PART			A TYPE OF COLOUR		
Prompted word	**1^{st}**	**2^{nd}**	**Prompted word**	**1^{st}**	**2^{nd}**
MUSIC	Head	Legs	**LENS**	Blue	Green

The words that you are asked to free associate BY considering the above-mentioned criterion are as follow:

	A HUMAN BODY PART				A TYPE OF COLOUR		
	Prompted word	**1st**	**2nd**		**Prompted word**	**1st**	**2nd**
1	**HUG**			**21**	**DARK**		
2	**PAIN**			**22**	**HAIR**		
3	**FOOD**			**23**	**OCEAN**		
4	**BROKEN**			**24**	**DRESS**		
5	**LICK**			**25**	**CLOUD**		
6	**WEAPONS**			**26**	**GRASS**		
7	**DRAWER**			**27**	**SUN**		
8	**SIGHT**			**28**	**ROSE**		
9	**WASH**			**29**	**LEAF**		
10	**LISTEN**			**30**	**SNOW**		
11	**HAT**			**31**	**EVIL**		
12	**RUN**			**32**	**GORDON**		
13	**RIVER**			**33**	**SAD**		
14	**NEEDLE**			**34**	**ELEPHANT**		
15	**HELP**			**35**	**OLD**		
16	**TAIL**			**36**	**AMATEUR**		
17	**TABLE**			**37**	**TREE**		
18	**CORN**			**38**	**COMMUNIST**		
19	**BOTTLE**			**39**	**METAL**		
20	**LANGUAGE**			**40**	**BRITISH**		

Table 25

Percentage of Three Most Common Words Given as a First or Second Response to the Primed Cues with Following the Category-Restriction Criterion (Cues Implies the Meanings of Target Either A Human Body Part or A Type of Colour)

Cues implying a body part meaning of the targets

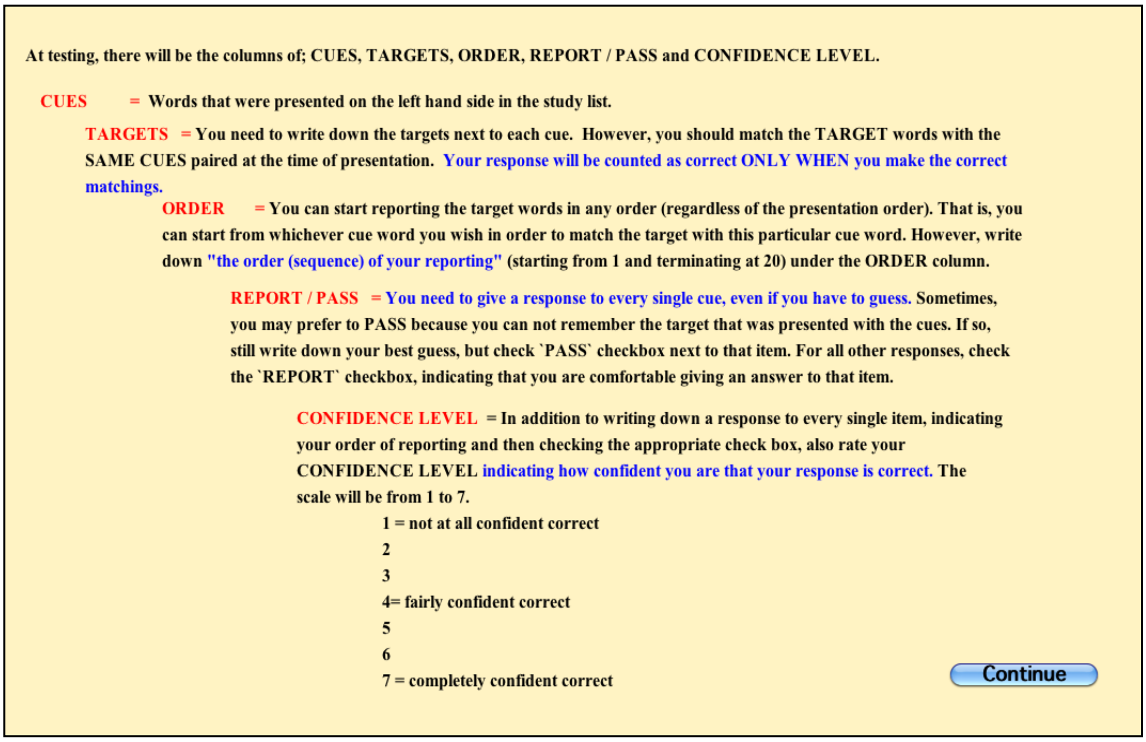
	Probable targets					Probable targets			
Cues	1st response	%	2nd response	%	Cues	1st response	%	2nd response	%
Hug	Arms	74.1	Chest	25.9	Weapons	Hand	25.9	Hands	11.1
	Arm	18.5	Torso	14.8		Hands	25.9	Head	11.1
	Hands	7.4	Body	7.4		Arms	14.8	Arm	7.4
Pain	Head	25.9	Arm	14.8	Drawer	Hand	37	Arm	25.9
	Heart	11.1	Head	14.8		Hands	14.8	Hands	14.8
	Arms	7.4	Leg	11.1		Legs	11.1	Arms	7.4
Listen	Ears	77.8	Head	29.6	Corn	Teeth	18.5	Mouth	33.3
	Ear	22.2	Eyes	18.5		Hands	14.8	Hands	11.1
	-	-	Brain	14.8		Mouth	14.8	Feet	7.4
Sight	Eyes	74.1	Head	37	Needle	Finger	29.6	Hand	22.2
	Eye	25.9	Brain	22.2		Eye	18.5	Arm	14.8
	-	-	Nose	11.1		Fingers	14.8	Finger	11.1
Wash	Hands	66.7	Face	37	Help	Hand	37	Hands	14.8
	Face	11.1	Feet	18.5		Hands	25.9	Head	11.1
	Hand	11.1	Arms	11.1		Mouth	14.8	Mouth	11.1
Hat	Head	100	Hair	33.3	Tail	Bottom	29.6	Legs	18.5
	-	-	Ears	18.5		Bum	18.5	Back	11.1
	-	-	Face	7.4		Back	14.8	Butt	7.4
Run	Feet	66.7	Feet	37	Table	Legs	37	Arms	14.8
	Foot	25.9	Legs	18.5		Leg	14.8	Hands	14.8
	Leg	3.7	Arms	7.4		Brown	7.4	Elbows	11.1
Food	Mouth	55.6	Stomach	37	River	Arms	44.4	Legs	40.7
	Stomach	40.7	Mouth	14.8		Feet	11.1	Feet	14.8
	Flesh	3.7	Tongue	11.1		Legs	11.1	Arm	7.4
Broken	Leg	33.3	Leg	22.2	Bottle	Mouth	40.7	Mouth	25.9
	Arm	14.8	Arm	18.5		Hand	22.2	Hand	14.8
	Bone	11.1	Heart	14.8		Hands	7.4	Hands	14.8
Lick	Tongue	77.8	Mouth	37	Language	Mouth	77.8	Brain	18.5
	Lips	14.8	Lips	14.8		Tongue	14.8	Ears	18.5
	Head	3.7	Tongue	11.1		Eyes	3.7	Head	14.8

Table 25 (*continued*).

Cues implying a type of colour meaning of the targets

	Probable targets					Probable targets			
Cues	1st response	%	2nd response	%	Cues	1st response	%	2nd response	%
Dark	Black	70.4	Navy	18.5	Evil	Black	44.4	Red	48.1
	Brown	14.8	Blue	14.8		Red	40.7	Black	37
	Blue	7.4	Black	11.1		Green	7.4	Grey	3.7
Hair	Black	33.3	Brown	44.4	Gordon	Brown	37	White	22.2
	Brown	29.6	Black	18.5		Green	22.2	Orange	14.8
	Blonde	14.8	Blonde	14.8		White	14.8	Yellow	14.8
Ocean	Blue	92.6	Green	48.1	Sad	Blue	40.7	Black	33.3
	Ears	3.7	White	14.8		Grey	14.8	Grey	25.9
	Green	3.7	Black	7.4		Black	11.1	Blue	14.8
Dress	Red	55.6	Black	29.6	Elephant	Grey	88.9	Brown	25.9
	Pink	11.1	Red	18.5		Arms	3.7	White	22.2
	White	11.1	Blue	7.4		Blue	3.7	Pink	14.8
Cloud	White	63	Grey	51.9	Old	Grey	48.1	Grey	22.2
	Blue	22.2	White	25.9		Black	11.1	Brown	11.1
	Grey	7.4	Blue	7.4		Brown	11.1	White	11.1
Grass	Green	96.3	Brown	51.9	Amateur	Blue	14.8	Blue	22.2
	Feet	3.7	Yellow	33.3		Brown	7.4	Brown	3.7
	-	-	Green	3.7		Black	3.7	-	-
Sun	Yellow	77.8	Orange	37	Tree	Green	63	Brown	63
	Orange	11.1	Red	18.5		Brown	33.3	Green	33.3
	Red	3.7	Yellow	14.8		Back	3.7	Palms	3.7
Rose	Red	48.1	Red	37	Communist	Red	92.6	Black	33.3
	Pink	44.4	Pink	22.2		Face	3.7	Yellow	25.9
	Nose	3.7	Green	11.1		Yellow	3.7	Brown	7.4
Leaf	Green	88.9	Brown	48.1	Metal	Silver	44.4	Black	18.5
	Ears	3.7	Red	11.1		Grey	33.3	Gold	18.5
	Red	3.7	Yellow	11.1		Black	7.4	Grey	18.5
Snow	White	92.6	Yellow	25.9	British	Red	40.7	Blue	29.6
	Lips	3.7	Brown	18.5		Blue	33.3	Red	29.6
	Yellow	3.7	Black	11.1		White	14.8	White	22.2

a) Instructions page for the cued-recall group



At testing, there will be the columns of; CUES, TARGETS, ORDER, REPORT / PASS and CONFIDENCE LEVEL.

CUES = Words that were presented on the left hand side in the study list.

TARGETS = You need to write down the targets next to each cue. However, you should match the TARGET words with the SAME CUES paired at the time of presentation. Your response will be counted as correct ONLY WHEN you make the correct matchings.

ORDER = You can start reporting the target words in any order (regardless of the presentation order). That is, you can start from whichever cue word you wish in order to match the target with this particular cue word. However, write down "the order (sequence) of your reporting" (starting from 1 and terminating at 20) under the ORDER column.

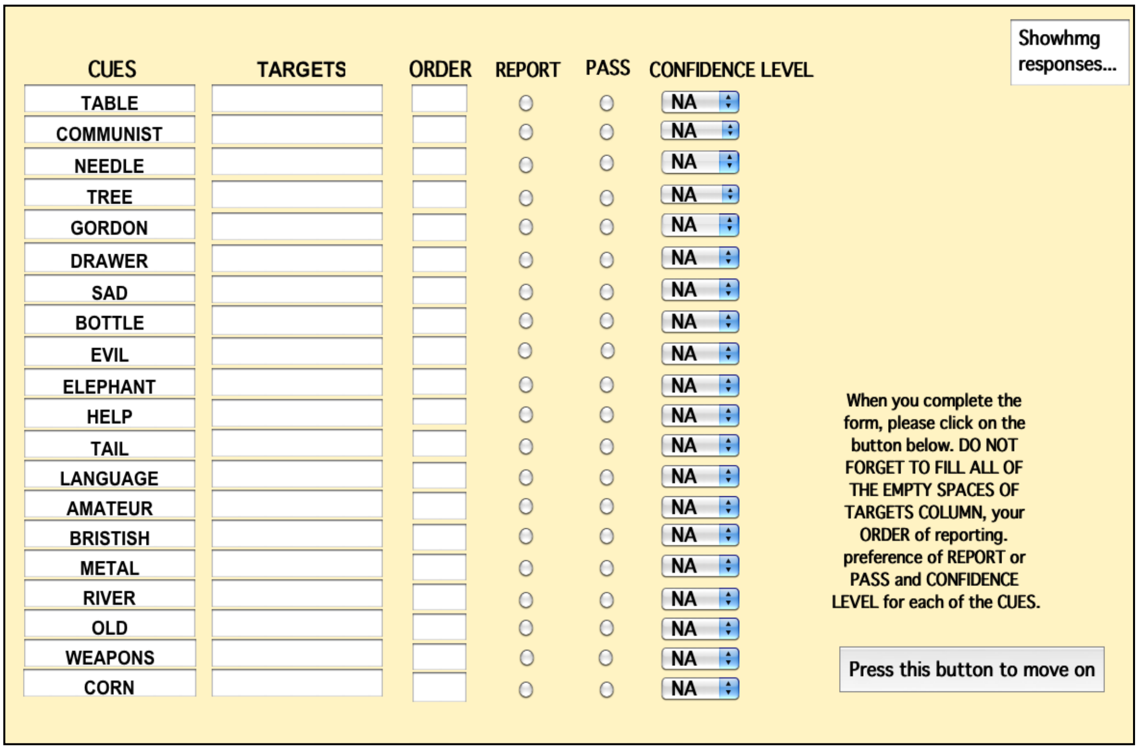
REPORT / PASS = You need to give a response to every single cue, even if you have to guess. Sometimes, you may prefer to PASS because you can not remember the target that was presented with the cues. If so, still write down your best guess, but check `PASS` checkbox next to that item. For all other responses, check the `REPORT` checkbox, indicating that you are comfortable giving an answer to that item.

CONFIDENCE LEVEL = In addition to writing down a response to every single item, indicating your order of reporting and then checking the appropriate check box, also rate your CONFIDENCE LEVEL indicating how confident you are that your response is correct. The scale will be from 1 to 7.

1 = not at all confident correct
2
3
4= fairly confident correct
5
6
7 = completely confident correct

Continue

b) Reporting page for the cued-recall group



CUES	TARGETS	ORDER	REPORT	PASS	CONFIDENCE LEVEL
TABLE					NA
COMMUNIST					NA
NEEDLE					NA
TREE					NA
GORDON					NA
DRAWER					NA
SAD					NA
BOTTLE					NA
EVIL					NA
ELEPHANT					NA
HELP					NA
TAIL					NA
LANGUAGE					NA
AMATEUR					NA
BRISTISH					NA
METAL					NA
RIVER					NA
OLD					NA
WEAPONS					NA
CORN					NA

Showhmg responses...

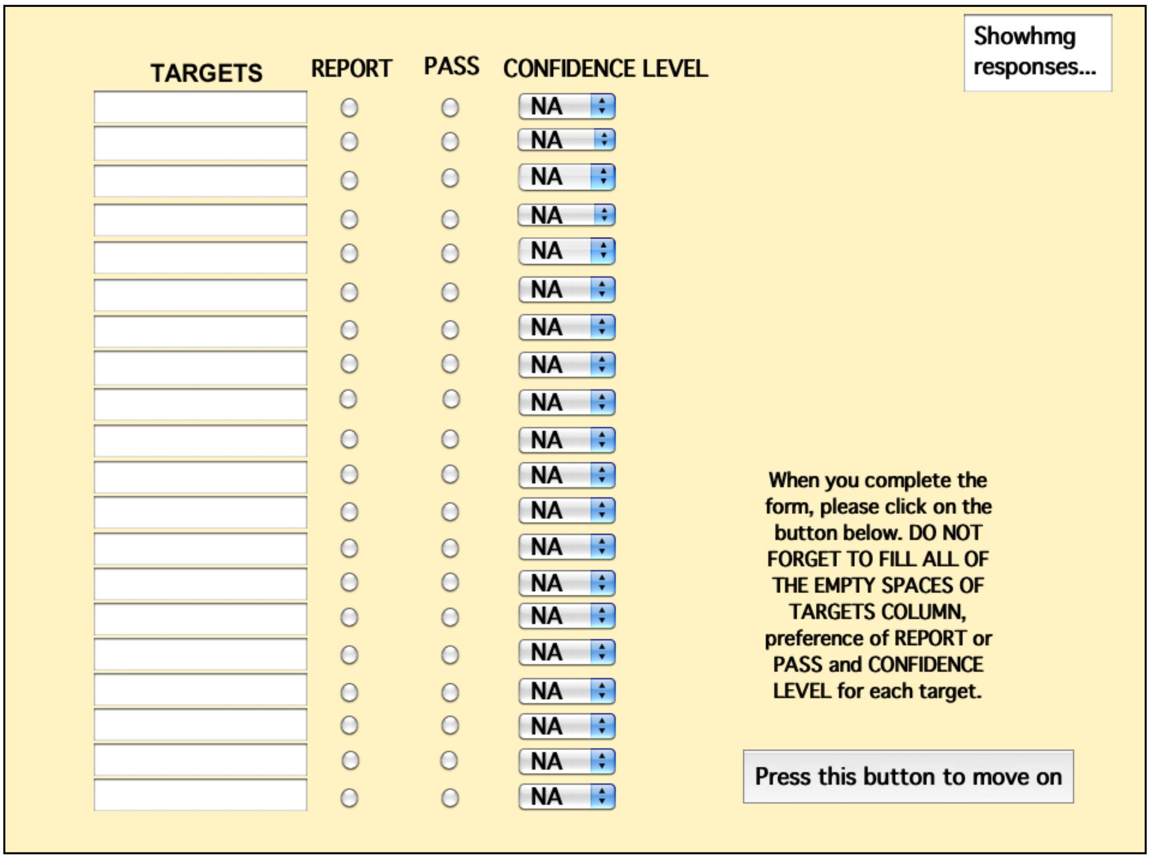
When you complete the form, please click on the button below. DO NOT FORGET TO FILL ALL OF THE EMPTY SPACES OF TARGETS COLUMN, your ORDER of reporting. preference of REPORT or PASS and CONFIDENCE LEVEL for each of the CUES.

Press this button to move on

Figure 21. The instructions page for the cued-recall group (a) and the pages of reporting for the cued-recall (b) and the uncued-recall groups (c) in Experiment 6.

Figure 21 (continued).

c) Reporting page for the uncued-recall group



Debriefing Statement

The aim of the research was to investigate the effects of inter-target association level on memory and metamemory performance. It was expected that when targets were matched with the cues that they imply a particular meaning with other targets, the memory performance would increase, however, monitoring performance of these targets would be lower compared to unrelated targets. Further, when participants use a shallow encoding strategy (repetition) they would have higher memory performance when targets were inter-related to the highest extend, however, their monitoring performance would be lower for the inter-related targets compared to the unrelated ones. However, when they used a deeper encoding strategy (interactive imagery), this dissociation would disappear. Your data will help our understanding of how the memory and metamemory performance are affected by the inter-target association level. The experiment did not use any deception. You may have a copy of this summary and if you wish you may also learn the summary of research findings with the following contact information.

If you have any further questions please contact me, Mehmet Akif Guzel, by the following e-mail address: **mag4v07@soton.ac.uk.**
Thank you for your participation in this research.

Signature ________________________________ Date __________________

Name Mehmet Akif Guzel

I understand that if I have questions about my rights as a participant in this research, or if I feel that I have been placed at risk, I can contact the Chair of the Ethics Committee, Department of Psychology, University of Southampton, Southampton, SO17 1BJ.
Phone: (023) 8059 5578

References

Abdi, H. (2007). Signal detection theory. In Salkind, N. (Ed.), *Encyclopedia of Measurement and Statistics,*(pp 886-889). CA: Sage Thousand Oaks. Retrieved from http://www.utdallas.edu/~herve/Abdi-SDT2007-pretty.pdf

Allwood, C. M., Granhag, P. A., & Jonsson, A-C (2006). Child witnesses' metamemory realism. *Scandinavian Journal of Psychology, 47*(6), 461-470. doi: 10.1111/1467-9450.00293

Allwood, C. M., Jonsson, A-C, & Granhag, P. A. (2005). The Effects of Source and Type of Feedback on Child Witnesses' Metamemory Accuracy. *Applied Cognitive Psychology, 19*(3), 3311-344. doi:10.1002/acp.1071

Anderson, J. R. (1983). A spreading activation theory of memory. *Journal of Verbal Learning and Verbal Behavior, 22,* 261-295. doi:10.1016/S0022-5371(83)90201-3

Anderson, J. R. & Bower, G. H. (1972). Recognition and Retrieval Processes in Free Recall. *Psychological Review, 79*(2), 97-123. doi:10.1037/h0033773

Antrobus, S. J. & Singer, J. L. (1964). Visual Signal Detection as a Function of Sequential Variability of Simultaneous Speech. *Journal of Experimental Psychology, 68*(6), 603-610. doi: 10.1037/h0043138

Atkinson, R.C. & Shiffrin, R.M. (1968). Human memory: A proposed system and its control processes. In K.W. Spence & J.T. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 2), (pp 742-775). New York: Academic. doi:10.1080/027249896392784

Bahrick, H. P. (1969). Measurement of memory by prompted recall. *Journal of Experimental Psychology, 79*(2), 213-219. doi:10.1037/h0026935

Bahrick, H. P. (1970). Two-phase model for prompted recall. *Psychological Review, 77*(3), 215-222. doi:10.1037/h0029099

Baker, L. & Santa, J. L. (1977). Context, Integration, and Retrieval. *Memory and Cognition, 5*(3), 308-314. doi:10.3758/BF03197575

Banks, W. P. (1970). Signal Detection Theory and Human Memory. *Psychological Bulletin, 74*(2), 81-99. doi:10.1037/h0029531

Bartling, C. A. & Thomson, C. P. (1977). Encoding Specificity: Retrieval Asymmetry in the Recognition Failure Paradigm. *Journal of Experimental Psychology: Human, Learning and Memory, 3*(6), 690-700. doi:10.1037//0278-7393.3.6.690

Battig, W. F. (1966). Evidence for coding processes “rote” paired-associate learning. *Journal of Verbal Learning and Verbal Behaviour*, *5,* 177-181. doi:10.1016/S0022-5371(66)80013-0

Battig, W. F. & Montague, W. (1969). Category Norms for verbal items in 56 categories: A Replication and Extension of the Connecticut Category Norms. *Journal of Experimental Psychology, 80*(3), 1- 46. doi:10.1037/h0027577

Benjamin, A. S. (2003). Predicting and postdicting the effects of word frequency on memory. *Memory & Cognition, 31*, 297– 305. doi:10.3758/BF03194388

Benjamin, A. S., & Bawa, S. (2004). Distractor plausibility and criterion placement in recognition. *Journal of Memory and Language, 51*, 159-172. doi:10.1016/j.jml.2004.04.001

Bilodeau, E. A. & Blick, K. A. (1965). Courses of misrecall over long-term retention intervals as related to strength of pre-experimental habits of word association. *Psychological Reports, 16,* 1173-1192.

Bousfield, W. A. (1953). The occurrence of clustering in the recall of randomly arranged associates. *Journal of General Psychology, 49*, 229-240. doi:10.1080/00221309.1953.9710088

Bower, G. H. (1980). Mood and Memory. *American Psychologist, 36*(2), 129-148. doi:10.1037/0003-066X.36.2.129

Bower, G. H. (2000). A Brief History of Memory Research. In Tulving, E. & Craik, F. I. M. (Eds). *The Oxford Handbook of Memory* (pp. 3-32). Oxford University Press, New York: USA.

Brown, S. C., & Craik, F. I. M. (2000). Encoding and retrieval of information. In E. Tulving & F.I.M. Craik (Eds.), *The Oxford handbook of memory* (pp. 93–107). New York: Oxford University Press.

Brown, S. C., Conover, J. N., Flores, L. M., & Goodman, K. M. (1991). Clustering and recall: Do high clusterers recall more than low clusterers because of clustering? *Journal of Experimental Psychology: Learning, Memory, & Cognition, 17*, 710-721. doi:10.1037//0278-7393.17.4.710

Calkins, M. W. (1895). Recognition and association. *Psychological Review, 2*(1), 94-95. doi:10.1037/h0069409

Chiarello, C., Liu, S., Quan, N., & Shears, C. (2000). Imagebility and word recognition in the left and right visual fields: A signal detection analysis. *Brain and Cognition, 43*(1-3), 90-94. doi:10.1006/brcg.1999.1133

Ciranni, M.A. & Shimamura, A.P. (1999) Retrieval- induced forgetting in episodic memory. J Exp Psychol Learn Mem Cogn 25, 1403-1414.

Clarke, F. R., Birdsall, T. G., & Tanner, W. P. Jr. (1959). Two types of ROC curves and definitions of parameters. *Journal of the Acoustical Society of America, 31*, 629-630. doi:10.1121/1.1907764

Cofer, C. N. (1967). Conditions for the use of verbal associations. *Psychological Bulletin, 68,* 1-12. doi:10.1037/h0024713

Cofer, C. N., Bruce, D. R., & Reicher, G. M. (1966). Clustering in free recall as a function of certain methodological variables. *Journal of Experimental Psychology, 71*, 858-866. doi:10.1037/h0023217

Cohen, B. H. (1966). Some or none characteristics of coding behavior. *Journal of Verbal Learning and Verbal Behavior, 5*, 182–187. doi:10.1016/S0022-5371(66)80014-2

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

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

Craik, F. I. M. & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General, 104*, 268-294. doi:10.1037//0096-3445.104.3.268

Crowder, R. G., & Greene, R. L. (2000). In E. Tulving & F. I. M. Craik (Eds.), *The Oxford Handbook of Memory* (pp. 149 - 162). New York: Oxford University Press.

De Vito, C. (1975). Encoding specificity and integration of word pairs. *Bulletin of the Psychonomic Society, 5*(3), 215-216.

Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology, 58*, 17-22.

Dewhurst, S.A., Bould, E., Knott, L.M., & Thorley, C. (2009). The roles of encoding and retrieval processes in associative and categorical memory illusions. *Journal of Memory & Language, 60*, 154-164. doi:10.1016/j.jml.2008.09.002

Dunlosky, J. T., & Hertzog, C. (1998a). Training programs to improve learning in later adulthood: Helping older adults educate themselves. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 249−275). Mahwah, NJ: Lawrence Erlbaum Associates.

Dunlosky, J., & Hertzog, C. (1998b). Aging and deficits in associative memory: What is the role of strategy production? *Psychology and Aging, 13*, 597−607. doi:10.1037//0882-7974.13.4.597

Dunlosky, J., & Nelson, T. O. (1992). Importance of the kind of cue for judgments of learning (JOL) and the delayed-JOL effect. *Memory & Cognition, 20*, 373-380. doi:10.3758/BF03210921

Dunlosky, J., & Nelson, T. O. (1994). Does the sensitivity of judgments of learning (JOLs) to the effects of various study activities depend on when the JOLs occur? *Journal of Memory and Language, 33*, 545-565. doi:10.1006/jmla.1994.1026

Dywan, J. & Bowers, K. (1983). *Science, 222*(4620), 184-185. doi:10.1126/science.6623071

Ebbinghaus, H. (1913). Memory: A contribution to experimental psychology. (H. A. Ruger & C. E. Bussenius, Trans.). New York: Teachers College Press. (Original work published 1885). doi:10.1037/10011-000

Epstein, M. L., Dupree, D.A., & Geonikowsky, L. A. (1979). Encoding specificity and contextual similarity. *Bulletin of the Psychonomic Society, 14*(3), 177-180.

Erdelyi, M. (1985). *Psychoanalysis: Freud's cognitive psychology*. New York, NY: W. H. Freeman.

Erdelyi, M. H., & Becker, J. (1974). Hypermnesia for pictures: Incremental memory for pictures but not words in multiple recall trials. *Cognitive Psychology, 6*, 159-171. doi:10.1016/0010-0285(74)90008-5

Erdelyi, M. H., Buschke, H., & Finkelstein, S. (1977). Hypermneasia for Socratic stimuli: The growth of recall for an internally generated memory list abstracted from a series of riddles. *Memory & Cognition, 5*(3), 283-286. doi:10.3758/BF03197571

Erdelyi, M. H., Finks, J., & Feigin-Pfau, M. B. (1989). The Effect of Response Bias on Recall Performance, with Some Observations on Processing Bias. *Journal of Experimental Psychology: General, 118*, 245-54. doi:10.1037//0096-3445.118.3.245

Flavell, J. H. (1979). Metacognition and Cognitive Monitoring: A New Area of Cognitive-Developmental Inquiry. *American Psychologist, 34* (10), 906– 911. doi:10.1037//0003-066X.34.10.906

Fry, E., Fountoukidis, D., & Polk, J. (1985). *The NEW Reading Teacher's Book of Lists*. Prentice-Hall, Inc. Englewood Cliffs, NJ. Retrieved from http://laurier.vsb.bc.ca/students/homgraphs.pdf

Gallo, D. A. (2010). False memories and fantastic beliefs: 15 years of the DRM illusion. Memory & Cognition, 37, 833-848. doi:10.3758/MC.38.7.833

Gallo, D. A., Weiss, J. A., & Schacter, D. L. (2004). Reducing false recognition with criterial recollection tests: Distinctiveness heuristic versus criterion shifts. *Journal of Memory and Language*, *51,* 473–493. doi:10.1016/j.jml.2004.06.002

Galvin, S. J., Podd, J. V., Drga, V., & Whitmore, J. (2003). Type 2 tasks in the theory of signal detectability: discrimination between correct and incorrect decisions. *Psychonomic Bulletin & Review, 10*, 843–876. doi:10.3758/BF03196546

Gardiner, J. M. & Tulving, E. (1980). Exceptions to Recognition Failure of Recallable Words. *Journal of Verbal Learning and Verbal Behaviour, 19,* 194-209. doi:10.1016/S0022-5371(80)90172-3

Goldsmith, M., & Koriat, A. (2008). The strategic regulation of memory accuracy and informativeness. In A. Benjamin & B. Ross (Eds.), *Psychology of Learning and Motivation, Advances in Research and Theory, Vol. 48*. New York: Academic Press. doi:10.1016/S0079-7421(07)48001-X

Gorman, A. M. (1961). Recognition memory for nouns as a function of abstractness and frequency. *Journal of Experimental Psychology, 6*, 123-129. doi:10.1037/h0040561

Granhag, P. A., Jonsson, A-C., & Allwood, C. M., (2004). The cognitive interview and its effect on witnesses' confidence. *Psychology, Crime & Law, 10*(1), 37-52. doi:10.1080/10683160210000030577

Green, (2011). In Merriam-Webster.com. Retrieved June 17, 2011, from http://www.merriam-webster.com/dictionary/green

Green, D.M. & Swets J.A. (1966). *Signal Detection Theory and Psychophysics*. New York: Wiley. doi:10.1016/0022-460X(67)90197-6

Guerin, S.A., & Miller, M. B. (2008). Semantic organization of study materials has opposite effects on recognition and recall. *Psychonomic Bulletin & Review, 15*(2), 302-308. doi:10.3758/PBR.15.2.302

Hacker, D. J., Bol, L., Horgan, D. D., & Rakow, E. A. (2000). Test prediction and performance in a classroom context. *Journal of Educational Psychology, 92*(1), 160-170. doi:10.1037//0022-0663.92.1.160

Hart, J. T. (1965). Memory and the Feeling-of-Knowing Experience. *Journal of Educational Psychology, 56*(4), 208–216. doi:10.1037/h0022263

Hart, J. T. (1966). Methodological Note on Feeling-of-Knowing Experiments. *Journal of Educational Psychology, 57*(6), 347–349. doi:10.1037/h0023915

Hertzog, C., Price, J., & Dunlosky, J. (2008a). How is knowledge generated about memory encoding strategy effectiveness? *Learning and Individual Differences, 18*, 430-445. doi:10.1016/j.lindif.2007.12.002

Higham, P. A. (2002). Strong cues are not necessarily weak: Thomson and Tulving (1970) and the encoding specificity principle revisited. *Memory & Cognition, 30*(1), 67- 80. doi:10.3758/BF03195266

Higham, P. A. (2011). Accuracy discrimination and type-2 signal detection theory: clarifications, extensions, and an analysis of bias. In, Higham, Philip A. & Leboe, Jason P. (eds.), *Constructions of Remembering and Metacognition. Essays in Honour of Bruce Whittlesea.* Basingstoke, GB, Palgrave Macmillan, 109-127.

Higham, P. A. & Guzel, M. A. (2011). Cued recall. In, Seel, Norbert (ed.) *Encyclopaedia of the Sciences and Learning.* London, Springer. (In Press). doi:10.1007/978-1-4419-1428-6

Higham, P. A. & Guzel, M. A. (2011). Cueing. In, Seel, Norbert (ed.) *Encyclopaedia of the Sciences and Learning.* London, GB, Springer. (In Press). doi:10.1007/978-1-4419-1428-6

Higham, P. A. & Tam, H. (2005). Generation failure: Estimating metacognition in cued recall. *Journal of Memory and Language, 52*, 595–617. doi:10.1016/j.jml.2005.01.015

Higham, P. A. & Tam, H. (2006). Release from generation failure: The role of study list structure. *Memory & Cognition, 34*(1), 148-157. doi:10.3758/BF03193394

Higham, P. A., & Arnold, M. M. (2007). How many questions should I answer? Using bias profiles to estimate optimal bias and maximum score on formula-scored tests. *European Journal of Cognitive Psychology, 19,* 718-742. doi:10.1080/09541440701326121

Higham, P. A., Luna, K. & Bloomfield, J. (2011). Trace-strength and source-monitoring accounts of accuracy and metacognitive resolution in the misinformation paradigm. *Applied Cognitive Psychology, 25*, 324-335. doi:10.1002/acp.1694

Howard, M. W. & Kahana, M. J. (1999). Contextual variability and serial position effects in free recall. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 25, 923–941. doi:10.1037//0278-7393.25.4.923

Howard, M. W. & Kahana, M. J. (2002). When does semantic similarity help episodic retrieval? *Journal of Memory and Language, 46*, 85-98. doi:10.1006/jmla.2001.2798

Hudson, R. L. & Austin, J. B. (1970). The effects of context and category name on the recall of categorized word lists. *Journal of Experimental Psychology, 86*(1), 43-47. doi:10.1037/h0029990

Hunt, R. R. & Ellis, H. C. (1999). *Fundamentals of Cognitive Psychology*. McGraw-Hill Humanities/Social Sciences, NV: USA.

Hunt, R. R., & Smith, R. E. (1996). Accessing the particular from the general: The power of distinctiveness in the context of organization. *Memory and Cognition, 24*, 217-225. doi:10.3758/BF03200882

Hunt, R.R. & Worthen, J. B. (Eds) (2006). *Distinctiveness and Memory*. NY: Oxford University Press, Inc.

Israel, L., & Schacter, D. L. (1997). Pictorial encoding reduces false recognition of semantic associates. *Psychonomic Bulletin & Review, 4*, 577–581. doi:10.3758/BF03214352

Jacoby, L. L. & Dallas, M. (1981). On the relationship between autobiographical and perceptual learning. *Journal of Experimental Psychology: General, 110*, 306-340. doi:10.1037//0096-3445.110.3.306

Jacoby, L. L. & Hollingshead, A. (1990). Toward a generate/recognize model of performance on direct and indirect tests of memory. *Journal of Memory and Language, 29,* 433–454. doi:10.1016/0749-596X(90)90065-8

Jacoby, L. L., Kelley, C. M. & Dywan, J. (1989). Memory attributions. In H. L. Roediger & F. I. M. Craik (Eds.), *Varieties of memory and consciousness: Essays in honour of Endel Tulving* (pp. 391-422). Hillsdale, NJ: Erlbaum.

Jenkins, J. J., Minx, W. D., & Russell, W. A. (1958). Associative clustering as a function of verbal association strength. *Psychol. Rep., 4*, 127- 136. doi:10.2466/PR0.4..127-136

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin*, 114, 3-28. doi:10.1037//0033-2909.114.1.3

Kato, T. (1985). Semantic-memory sources of episodic retrieval failure. *Memory & Cognition, 13*(5), 442-452. doi:10.3758/BF03198457

Keister, R. S. (1972). The effects of blocked versus randomized presentation orders on organization in paired-associate learning. *Journal of Verbal Learning and Verbal Behavior, 11*, 293-302. doi:10.1016/S0022-5371(72)80090-2

Kelley, C. M., & Sahakyan, L. (2003). Memory, monitoring, and control in the attainment of memory accuracy. *Journal of Memory & Language, 48*, 704-721. doi:10.1016/S0749-596X(02)00504-1

King, J. F., Zechmeister, E. B., & Shaughnessy, J. J. (1980). Judgments of knowing: The influence of retrieval practice. *American Journal of Psychology, 93*, 329-343. doi:10.2307/1422236

Kintsch, W. (1968). Recognition and Free Recall of Organize Lists. *Journal of Experimental Psychology, 78*(3), 481–487. doi:10.1037/h0026462

Kintsch, W. (1970). *Learning, memory, and conceptual processes.* New York: Wiley.

Klatzky, R. L. (1975). *Human Memory: Structures and Processes*. W. H. Freeman Company, USA.

Klatzky, R. L., & Erdelyi, M. H. (1985). The response criterion problem in tests of hypnosis and memory. *International Journal of Clinical and Experimental Hypnosis, 33,* 246-257. doi:10.1080/00207148508406653

Knott, L.M., & Dewhurst, S.A. (2007). The effects of divided attention at study and test on false recognition: A comparison of DRM and categorized lists. *Memory & Cognition, 35*, 1954-1965. doi:10.3758/BF03192928

Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General, 126*, 349–370. doi:10.1037/0096-3445.126.4.349

Koriat, A. & Goldsmith, M. (1994). Memory in naturalistic and laboratory contexts: distinguishing the accuracy-oriented and quantity-oriented approaches to memory assessment. *Journal of Experimental Psychology: General, 123*(3), 297-315. doi:10.1037//0096-3445.123.3.297

Koriat, A. & Goldsmith, M. (1996a). Memory as something that can be counted versus memory as something that can be counted on. In D. Hermann, C. McEvoy, C. Hertzog, P. Hertel, & M. Johnson (Eds.), *Basic and applied memory research: Practical applications* (Vol. 2, pp. 3-18). Hillsdale, NJ: Erlbaum.

Koriat, A. & Goldsmith, M. (1996b). Memory metaphors and the real-life/laboratory controversy: Correspondence versus storehouse conceptions of memory. *Behavioural & Brain Sciences, 19*, 167-228.

Koriat, A. & Goldsmith, M. (1996c). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review*, 103, 490-517. doi:10.1037//0033-295X.103.3.490

Koriat, A., Levy-Sadot, R. (2001). The combined contributions of the cue familiarity and accessibility heuristics to feeling of knowing. *Journal of Experimental Psychology: Learning, Memory and Cognition, 27*, 34-53. doi:10.1037//0278-7393.27.1.34

Koriat, A. & Melkman, R., (1987). Depth of processing and memory organization. *Psychological Research, 49*, 173-181. doi:10.1007/BF00308684

Kucera, H. & Francis, W. N. (1967). Computational Analysis of Present-day American English. Providence: Brown University press. Retrieved from http://www.psych.rl.ac.uk/MRC_Psych_Db_files/mrc2.html

Lauer, P. A. (1974). Encoding specificity in the cued and free recall of categorically and alphabetically organized words. *Bulletin of the Psychonomic Society, 4*(5), 496-498.

Lindsay, D. S., & Kelley, C. M. (1996). Creating illusions of familiarity in a cued recall remember/know paradigm. *Journal of Memory and Language*, 35, 197-211. doi:10.1006/jmla.1996.0011

Lockhart, R. S. (1969). Retrieval asymmetry and the criterion problem in cued recall. *Journal of Experimental Psychology, 81*(1), 192-194. doi: 10.1037/h0027466

Luek, M. Q., Mclauglin, J. P., & Cicala, G. A. (1971). Effects of blocking input and blocking of retrieval cues on free recall learning. *Journal of Experimental Psychology, 91*, 159-161. doi:10.1037/h0031842

MacMillan, N. A. & Creelman, C. D. (1991). *Detection Theory: A user's guide*. New York: Cambridge University.

MacMillan, N. A. & Creelman, C. D. (2004). *Detection Theory: A user's guide* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Macmillan, N. A. & Creelman, C. D. (2005). *Detection Theory: A User's Guide*. Lawrance Erlbaum Associates, USA.

Multhaup K. S. & Balota D. A. (1997). Generation effects and source memory in healthy older adults and adults with dementia of the Alzheimer type. *Neuropsychology, 11*, 382–391. doi:10.1037//0894-4105.11.3.382

Multhaup, K. S., De Leonardis, D. M., & Johnson (1999). Source memory and eyewitness suggestibility in older adults. *Journal of General Psychology, 126*, 74-84. doi:10.1080/00221309909595352

Mandler, G., Pearlstone, Z., & Koopmans, H. J. (1969). Effects of organization and semantic similarity on recall and recognition. *Journal of Verbal Learning and Verbal Behaviour, 8*, 410-423. doi:10.1016/S0022-5371(69)80134-9

Mantyla, T., & Nilsson, L. G. (1988). Cue distinctiveness and forgetting: effectiveness of self- generated retrieval cues in delayed recall. *Journal of Experimental Psychology: Learning Memory and Cognition,* 14, 502–509. doi:10.1037//0278-7393.14.3.502

Martin, E. (1975). Theoretical Notes: Generation-Recognition Theory and the Encoding Specificity Principle. *Psychological Review, 82*(2), 150–153. doi:10.1037/h0076779

May, R. B. & Tryk, H. E. (1970). Word Sequence, Word Frequency, and Free Recall. *Canadian Journal of Psychology, 24*(5), 299-304. doi:10.1037/h0082866

McEvoy, C. L., Nelson, D. L., Holley, P. E., & Stelnicki, G. S. (1992). Implicit Processing in the Cued Recall of Young and Old Adults. *Psychology and Aging, 7*(3), 401-408. doi:10.1037//0882-7974.7.3.401

Mclaughlin, B. (1965). "Intentional" and "Incidental" learning in human subjects: the role of instructions to learn and motivation. *Psychological Bulletin, 63*(8), 359-376. doi:10.1037/h0021759

Miller, R. J. & Leibowitz, H. W. (1976). A Signal Detection Analysis of Hypnotically Induced Narrowing of the Peripheral Visual Field. *Journal of Abnormal Psychology, 85*(5), 446-454. doi:10.1037//0021-843X.85.5.446

Moulin, C. J. A. (2002). Sense and sensitivity: Metacognition in Alzheimer's disease. In T.J. Perfect & B.L. Schwartz (Eds.), *Applied metacognition*. Cambridge, UK: Cambridge University Press.

Moulin, C. J. A., Perfect, T. J., & Jones, R. W. (2000a). The effects of repetition on allocation of study time and judgements of learning in Alzheimer's disease. *Neuropsychologia, 38*, 750–758. doi:10.1016/S0028-3932(99)00142-6

Moulin, C. J. A., Perfect, T. J., & Jones, R.W. (2000b). Global predictions of memory in Alzheimer's disease: Evidence for preserved metamemory monitoring. Aging, *Neuropsychology, and Cognition, 7*, 230–244. doi:10.1076/anec.7.4.230.796

Moulin, C. J. A., Perfect, T. J., & Jones, R. W. (2000c). Evidence for intact memory monitoring in Alzheimer's disease: Metamemory sensitivity at encoding. *Neuropsychologia, 38*, 1242–1250. doi:10.1016/S0028-3932(00)00037-3

Moulin, C. J. A., Perfect, T. J., Conway, M. A., North, A. S., Jones, R. W. & James, N. (2002). Retrieval-induced forgetting in Alzheimer's disease. *Neuropsychologia, 40*(7), 862-867. doi:10.1016/S0028-3932(01)00168-3

Muter, P. (1984). Recognition and Recall of Words With a Single Meaning. *Journal of Experimental Psychology: Learning, Memory and Cognition, 10*(2), 198-202. doi:10.1037//0278-7393.10.2.198

Nairne, J. S. (2002). The myth of the encoding-retrieval match. *Memory, 10*(5/6), 389–395. doi:10.1080/09658210244000216

Nelson, D. L. & McEvoy, C. L. (1979). Encoding context and set size. *Journal of Experimental Psychology: Human Learning and Memory, 5*(3), 292-314. doi:10.1037//0278-7393.5.3.292

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1990). Encoding Context and Retrieval Conditions as Determinants of the Effects of Natural Category Size. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*(1), 31-41. doi:10.1037//0278-7393.16.1.31

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1998). The University of South Florida word association, rhyme, and word fragment norms. Retrieved from http://www.usf.edu/FreeAssociation/.

Nelson, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Bulletin, 95*, 109–133. doi:10.1037//0033-2909.95.1.109

Nelson, T. O., Gerler, D., & Narens, L. (1984). Accuracy of feeling of knowing judgements for predicting perceptual identification and relearning. *Journal of Experimental Psychology: General, 113*, 282-300. doi:10.1037//0096-3445.113.2.282

Nelson, T. O. & Narens, L. (1994). Why investigate metacognition? In J. Metcalfe, and A. P. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp.1-25). Cambridge, MA: MIT Press.

Nelson, T. O. & Leonesio, R. J. (1988). Allocation of Self-Paced Study Time and the "Labor-in-Vain Effect". *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*(4), 676-686. doi:10.1037//0278-7393.14.4.676

Newman, S. E., Cooper, M. H., Parker, K. O., Sidden, J. A., Gonder, L. A., Moorefield, K. M., & Nelson, P. A. (1982). Some tests of the encoding specificity and semantic integration hypotheses. *American Journal of Psychology, 95*(1), *103-123.* doi:10.2307/1422662

Pavio, A. (1978). Comparisons of mental clocks. *Journal of Experimental Psychology: Human Perception and Performance,* 4, 61-71. doi:10.1037//0096-1523.4.1.61

Pellegrino, J. W. & Salzberg, P. M. (1974). Encoding Specificity in Cued Recall and Context Recognition. *Journal of Experimental Psychology: Human, Learning and Memory, 104*(3), 261-270. doi:10.1037//0278-7393.1.3.261

Pellegrino, J. W. & Salzberg, P. M. (1975). Encoding specificity in associative processing tasks. *Journal of Experimental Psychology: Human, Learning and Memory, 1*(5), 538-548. doi:10.1037//0278-7393.1.5.538

Perfect, T. J. & Schwartz, B. L. (2002). *Applied Metacognition*. Cambridge University Press, Cambridge: UK.

Polyn S. M., Norman K. A., & Kahana M. J. (2009b). A context maintenance and retrieval model of organizational processes in free recall. *Psychological Review*, 116(1), 129-156. doi:10.1037/a0014420

Polyn, S. M., Norman, K. A., & Kahana, M. J. (2009a). Task context and organization in free recall. *Neuropsychologia, 47*, 2158-63. doi:10.1016/j.neuropsychologia.2009.02.013

Posner, M. O., & Keele, S. W. (1970). Retention of abstract ideas. *Journal of Experimental Psychology, 83*, 304-308. doi:10.1037/h0028558

Postman, L. (1963). One trial learning. In C. F. Cofer & B. S. Musgrave (Eds.), *Verbal behaviour and learning* (pp. 295-332). New York: McGraw-Hill.

Puff, C. R. (1970). Role of clustering in free recall. *Journal of Experimental Psychology, 86*, 384-386. doi:10.1037/h0030189

Puff, C. R., Murphy, M. D., & Ferrara, R. A. (1977). Further evidence about the role of clustering in free recall. *Journal of Experimental Psychology: Human, Learning, and Memory, 3*(6), 742-753. doi:10.1037//0278-7393.3.6.742

Quillian, M. R. (1968). Semantic memory. In M. Minsky (ed), *Semantic information processing*. Cambridge, Mass.: M.I.T. Press.

Quillian, M. R. (1969). The teachable language comprehender: A simulation program and theory of language. *Communications of the ACM, 12*, 459-476. doi:10.1145/363196.363214

Reder, L. M., Anderson, J. R., & Bjork, R. A. (1974). A Semantic Interpretation of Encoding Specificity. *Journal of Experimental Psychology, 102* (4), 648–656. doi:10.1037/h0036115

Reyna, V. F., & Brainerd, C. J. (1995). Fuzzy-trace theory: An interim synthesis. *Learning and Individual Differences, 7*, 1–75. doi:10.1016/1041-6080(95)90031-4

Roberts, W. T. & Higham, P. A. (2002). Selecting accurate statements from the cognitive interview using confidence ratings. *Journal of Experimental Psychology: Applied, 8*(1), 33-43. doi:10.1037//1076-898X.8.1.33

Roeder, C. M. (2006). Developmental progression in children's strategic memory regulation. *Swiss Journal of Psychology (Schweizerische Zeitschrift für Psychologie), 65*(3), 193-200. doi:10.1024/1421-0185.65.3.193

Roediger, H. L. & Adelson, B. (1980). Semantic specificity in cued recall. *Memory and Cognition, 8*(1), 65-74. doi:10.3758/BF03197553

Roediger, H. L. & Payne, D. G. (1983). Superiority of Free Recall to Cued Recall with "Strong" Cues. *Psychological Research, 45*, 275-286. doi:10.1007/BF00308707

Roediger, H. L., & Payne, D. G. (1982). Hypermnesia: The role of repeated testing. *Journal of Experimental Psychology: Learning, Memory and Cognition, 8*, 66-72. doi:10.1037//0278-7393.8.1.66

Roediger, H. L., & Payne, D. G. (1985). Recall criterion does not affect recall level or hypermnesia: A puzzle for generate/recognize theories. *Memory & Cognition, 13*, 1-7. doi:10.3758/BF03198437

Roediger, H. L., III, & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 21, 803–814. doi:10.1037//0278-7393.21.4.803

Roediger, H. L., III, & Thorpe, L. A. (1978). The role of recall time in producing hypermnesia. *Memory & Cognition, 6*, 296–305.

Roediger, H. L., III, Watson, J. M., McDermott, K. B., & Gallo, D. A. (2001). Factors that determine false recall: A multiple regression analysis. *Psychonomic Bulletin & Review, 8*, 385-407. doi:10.3758/BF03196177

Roediger, H. L., Srivinas, K., & Waddil (1989). How Much Does Guessing Influence Recall? Comment on Erdelyi, Finks, and Feigin-Pfau. *Journal of Experimental Psychology: General, 118*(3), 255-257. doi:10.1037//0096-3445.118.3.255

Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General, 104*, 192-233. doi:10.1037//0096-3445.104.3.192

Rosch, E. (1978). Principles of categorization. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization*. Hillsdale, NJ: Erlbaum. Reprinted in: Margolis, E. & Laurence, S. (Eds.) (1999), *Concepts: Core readings.* Cambridge, MA: MIT Press.

Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology, 7*, 573-605. doi:10.1016/0010-0285(75)90024-9

Runquist, P. A. (1970). Clustering in free recall following paired associate learning. *Journal of Experimental Psychology, 83*, 260-265. doi:10.1037/h0028546

Salzberg, P. M. (1975). On the generality of encoding specificity. *Journal of Experimental Psychology: Human Learning and Memory 2*(5), 586-596. doi:10.1037//0278-7393.2.5.586

Salzberg, P. M. & Pellegrino, J. W. (1974). The generation and recognition components of encoding specificity. *Bulletin of the Psychonomic Society, 4*(1), 9-11.

Sanabria, D., Spence, C., & Soto-Faraco, S. (2008). Perceptual and decisional contributions to audiovisual interactions in the perception of apparent motion: A signal detection study. *Cognition, 102*, 299–310. doi:10.1016/j.cognition.2006.01.003

Santa, J. L. & Lamwers, L. L. (1974). Encoding specificity: fact or articact. *Journal of Verbal Learning & Verbal Behaviour, 13*(4), 412-423. doi:10.1016/S0022-5371(74)80019-8

Santa, J. L. & Lamwers, L. L. (1976). Where Does the Confusion Lie?: Comments on the Wiseman and Tulving Paper. *Journal of Verbal Learning and Verbal Behaviour, 15*, 53-57. doi:10.1016/S0022-5371(76)90006-2

Sauer, J. D., Brewer, N., & Weber, N. (2008). Multiple Confidence Estimates as Indices of Eyewitness Memory. *Journal of Experimental Psychology: General, 137*(3), 528–547. doi:10.1037/a0012712

Schacter, D. L., Israel, L., & Racine, C. (1999). Suppressing false recognition in younger and older adults: The distinctiveness heuristic. *Journal of Memory & Language, 40*, 1-24. doi:10.1006/jmla.1998.2611

Schulman, A. I. (1971). Recognition Memory for Targets from a Scanned Word List. *British Journal of Psychology, 62*, 335–346. doi:10.1111/j.2044-8295.1971.tb02044.x

Schwartz, S. (1975). Encoding specificity and recognition memory for words. *Bulletin of the Psychonomic Society, 6*(3), 279-281.

Shiffrin, R. M. & **Atkinson**, R. C. (1969). Storage and retrieval processes in long-term memory. *Psychological Review, 76*(2), 179-193. doi:10.1037/h0027277

Slamecka, N. J. (1968). An examination of trace storage in free recall. *Journal of Experimental Psychology, 76*, 504–513. doi:10.1037/h0025695

Slamecka, N. J. (1972). The question of associative growth in the learning of categorized materials. *Journal of Verbal Learning and Verbal Behavior, 11*, 324–332. doi:10.1016/S0022-5371(72)80094-X

Slamecka, N. J. (1985). Ebbinghaus: Some associations. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*, 414-435. doi:10.1037//0278-7393.11.3.414

Slamecka, N., & Graf, P. (1978). The generation effect: delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning and Memory, 14*, 592-604. doi:10.1016/S0022-5371(83)90112-3

Smith, E. E., Shoben, E. J. & Rips, L. J. (1974). Structure and process in semantic memory: A featural model for semantic decisions. *Psychological Review, 1*, 214-241. doi:10.1037/h0036351

Smith, S. M., Gerkens, D. R., Pierce, B. H., & Choi, H. (2002). The roles of associative responses at study and semantically guided recollection at test in false memory: The Kirkpatrick and Deese hypotheses. *Journal of Memory & Language, 47*, 436-447. doi: 10.1016/S0749-596X(02)00012-8

Stein, B. (1978). Depth of processing reexamined: The effects of the precision of encoding and test appropriateness. *Journal of Verbal Learning and Verbal Behavior, 17,* 165–174. doi:10.1016/S0022-5371(78)90128-7

Thompson, C. P., Hamlin, V. J., & Roenker, D. L. (1972). A comment on the role of clustering in free recall. *Journal of Experimental Psychology, 94*, 108–109. doi:10.1037/h0032797

Thomson, D. M. & Tulving, E. (1970). Associative Encoding Retrieval: Weak and Strong Cues. *Journal of Experimental Psychology, 86*(2), 255–262. 10.1037/h0029997

Tresselt, M. E. & Mayzner, M. S. (1960). A study of incidental learning. *Journal of Psychology, 50*, 339-47. doi:10.1080/00223980.1960.9916451

Tulving, E. (1962). Subjective organization in free recall of “unrelated” words. *Psychological Review, 69*, 344-354. doi:10.1037/h0043150

Tulving, E. (1966). Subjective Organization and Effects of Repetition in Multi-Trial Free-Recall Learning. *Journal of Verbal Learning and Verbal Behavior, 5*, 193-197. doi:10.1016/S0022-5371(66)80016-6

Tulving, E. (1974). Recall and Recognition of Semantically Encoded Words. *Journal of Experimental Psychology, 102*(5), 778–787. doi:10.1037/h0036383

Tulving, E. (1982). Synergistic ecphory in recall and recognition. *Canadian Journal of Psychology*, *36*, 130-147. doi:10.1037/h0080641

Tulving, E. & Osler, S. (1968). Effectiveness of retrieval cues in memory for words. *Journal of Experimental Psychology, 77*(4), 593-601.

Tulving, E. & Thomson, D. M. (1971). Retrieval Processes in Recognition Memory: Effects of Associative Context. *Journal of Experimental Psychology, 87*(1), 116–124. doi:10.1037/h0030186

Tulving, E. & Thomson, D. M. (1973). Encoding Specificity and Retrieval Processes in Episodic Memory. *Psychological Review, 80*(5), 352-373. doi:10.1037/h0020071

Tulving, E. & Watkins, O. C. (1977). Recognition failure of words with a single meaning. *Memory and Cognition, 5*(5), 513-522. doi:10.3758/BF03197394

Tulving, E., & Craik, F. I. M. (2000). *The Oxford Handbook of Memory*. Oxford: Oxford University Press.

Tulving, E., & Osler, S. (1968). Effectiveness of retrieval cues in memory for words. *Journal of Experimental Psychology, 77*, 593-601. doi:10.1037/h0026069

Tulving, E., & Pearlstone, Z. (1966). Availability versus accessibility of information in memory for words. *Journal of Verbal Learning and Verbal Behavior, 5*, 381-391. doi:10.1016/S0022-5371(66)80048-8

Van Overschelde, J. P. (2002). The Influence of Word Frequency on Recency Effects in Directed Free Recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*(4), 611–615. doi:10.1037//0278-7393.28.4.611

Van Overschelde, J. P., Rawson, K. A., & Dunlosky, J. (2004). Category norms: An updated and expanded version of the Battig and Montague (1969) norms. *Journal of Memory and Language, 50*, 289–335. doi:10.1016/j.jml.2003.10.003

Watkins, J. & Gardiner, J. M. (1979). An Appreciation of Generate-Recognize Theory of Recall. *Journal of Verbal Learning and Verbal Behaviour, 18,* 687–704. doi:10.1016/S0022-5371(79)90397-9

Wickens, T. D. (2002). *Elementary Signal Detection*. New York, USA: Oxford University Press.

Wiseman, S. & Tulving, E. (1976). Encoding Specificity: Relation Between Recall Superiority and Recognition Failure. *Journal of Experimental Psychology: Human Learning and Memory, 2*(4), 349-361. doi:10.1037//0278-7393.2.4.34