Title	二重レベルのメタ認知と学習戦略選択のメカニズム：認知的オフローディングの事例
Author(s)	馬, 遠
Citation	
Issue Date	2025-03
Type	Thesis or Dissertation
Text version	ETD
URL	http://hdl.handle.net/10119/19913
Rights	
Description	Supervisor: 藤波 努, 先端科学技術研究科, 博士

Doctoral Dissertation

The Mechanism between Two-layer Metacognition and Strategy Choice for Learning: The Case of Cognitive Offloading

Yuan Ma

Supervisor: Tsutomu Fujinami

Graduate School of Advanced Science and Technology

Japan Advanced Institute of Science and Technology

[Knowledge Science]

March 2025

Supervised by:

Prof Dr Tsutomu Fujinami

Reviewed by:

Prof Dr Dam Hieu Chi

Prof Dr Mizumoto Masaharu

Prof Dr Yoshioka Hidekazu

Prof Dr Kojima Haruyuki

Abstract

Individuals may assess their memory performance and, in turn, rely on external aids to ensure that information remains accessible for future retrieval. For example, an individual might use a password manager if they consider a newly created password challenging to remember. However, perceived memory performance often differs from actual memory performance, making it challenging to predict when external assistance is cognitively necessary. Related research defines self-assessment of cognitive performance as metacognition (e.g., perceived memory performance) and the use of external resources to support cognitive functioning as cognitive offloading (e.g., using the password manager). Accordingly, the unpredictability of cognitive offloading based on perceived memory performance complicates understanding the mechanisms of metacognition involved in deciding whether to employ cognitive offloading. This dissertation introduced confidence in the previous assessment of perceived memory performance as a second-order metacognitive judgement (SOJ) to address this issue, considering the initial assessment as a first-order metacognitive judgement (FOJ). This research examined the following question: "How do individuals decide to employ cognitive offloading based on two-layer self-assessment in learning tasks?"

This research used 48 English paired associates (e.g., ABILITY-CAPABILITY) as learning tasks, incorporating a procedure with Learning, Retention, and Test sessions. In the Learning session, participants were instructed to learn each associate and then estimate their performance to recall the target item (e.g., CAPABILITY) when presented with the cue item (e.g., ABILITY) in a later test. Subsequently, their confidence in the correctness of FOJ was elicited as their SOJ during the trial. Afterwards, the participants chose a learning strategy for each associate: write it down on paper (employing cognitive offloading) or remember it mentally (not employing cognitive offloading). In the Retention session, the participants completed simple mathematical problems. In the Test session, participants were required to freely recall the target items for all paired associates (inputting via a keyboard) when presented with the cues. The learning tasks were administered online, with participants recruited remotely from the United States.

This research comprised three studies: Studies 1 and 2 served as pilot studies, and Study 3 was the main study. Study 1 verified the suitability of the procedure and learning materials for subsequent studies. Study 2 determined a scoring method for strategy choices to reduce the overuse of cognitive offloading. Study 3 provided evidence to address the research question, finding that (1) FOJs partially predict the selection of cognitive offloading and that (2) SOJs influence the regulation of cognitive offloading choices. The increased consistency with actual memory performance in metacognition may result from confidence regulating the strategy choice relative to the previous FOJ to the opposite option if the SOJ falls below a certain threshold. For example, shallow confidence may lead to switching from relying on memory to offloading information, even if the target was initially perceived as memorable.

This research initially explored how SOJs influence learning strategies, contributing to knowledge science by highlighting metaknowledge as a novel aspect of knowledge creation. On another note, this work provides insights into unresolved issues in cognitive offloading and related research fields that employ metacognitive judgements as a methodology. Additionally, this work has practical implications for educational contexts where individuals interact with information in their environment. Future research could expand on these findings by incorporating a broader range of learning strategies beyond cognitive offloading, exploring diverse methods for eliciting confidence levels, and extending the current findings with various alternative materials.

Keywords

Metacognition, memory, cognitive offloading, metacognitive judgement, monitoring, control

Table of Contents

List of Figures

List of Tables

List of Abbreviations

FOJ: First-order Metacognitive Judgement

SOJ: Second-order Metacognitive Judgement

MTurk: Amazon Mechanical Turk

HIT: Human Intelligence Task

USD: United States Dollar

M: Mean

SD: Standard Deviation

AD: Anderson-Darling

CI: Confidential Interval

ANOVA: Analysis of Variance

VAR: Variance

AICc: Akaike’s Corrected Information Criterion

BIC: Bayesian Information Criterion

STEM: Science, Technology, Engineering, and Mathematics

Chapter 1 Introduction

Overview

Individuals often offload information through physical actions when relying solely on memory becomes burdensome. For example, they may prefer using a password manager for a newly registered complex password to avoid the mental effort required to remember it for future logins. Individuals' perceived memory performance based on their self-assessments significantly influences this decision. Additionally, their confidence levels in these assessments provide further insights. This dual consideration makes using physical actions to reduce mental workload more predictable than relying solely on self-assessment of perceived memory performance. This dissertation examined the interaction of these two factors in strategy choice for recalling information from a cognitive psychology perspective on the monitoring-control relationship. Research in this field can potentially enhance the allocation of external resources, thereby improving learning gain and working efficiency. This chapter begins with a brief history of scientific research on human memory, followed by a theoretical overview of using physical actions as an alternative to memory function (i.e., cognitive offloading) and its connection to self-assessments on memory performance (i.e., metacognition). Finally, this chapter presents the dissertation's research question and main hypotheses.

Portions of this dissertation, including descriptions of the theoretical framework, methodology, results, alternative interpretations, limitations, conclusions, and impacts, are included in the article "The influence of second-order metacognitive judgements on cognitive offloading within the monitoring-control relationship," published in *Discover Psychology* under a CC BY 4.0 licence (Ma & Fujinami, 2024).

1.1 Memory

Memory is the cognitive ability and process through which information, experiences, and knowledge are encoded, stored, and retrieved as needed. It involves the brain's capacity to acquire, organise, and store information for a few milliseconds to a lifetime. Memory encompasses various mental functions, from recalling specific facts and events to recognising familiar connections. Memory is fundamental to human cognition, learning, and everyday functioning.

In the beginning of memory research, the questions were naturally focused on the temporal dimension – specifically, how long information can be effectively retained. The complex process of human memory involves three distinct stages: *sensory*, *short-term*, and *long-term*. Sensory memory is the initial stage of memory processing, serving as a rapid repository for sensory information that momentarily retains raw sensory data. This brief storage of sensory input – such as what we see (Sperling, 1960), hear (Darwin et al., 1972), or feel (Bliss et al., 1966) – enables us to perceive and interpret the world around us. Sensory memory has the shortest duration of the three memory stages, typically lasting only milliseconds to a few seconds. Short-term memory, the second stage in the memory process, temporarily retains and processes information received from sensory memory or retrieved from long-term memory. It lasts longer than does sensory memory, typically holding information for several seconds to a few minutes. Short-term memory is often described as a workspace where information is actively manipulated, making it available for immediate cognitive tasks. Also, related research terms the function of manipulating information as short-term memory included by *working* memory if for inferencing (Baddeley, 1992), or *micro-term* memory in addition to short-term memory if for regulating skill execution (Fujinami & Hidaka, 2019). Long-term memory, the final stage of the memory process, involves the relatively permanent storage of information consolidated from short-term memory (McGaugh, 2000). Information in long-term memory can persist for days, weeks, months, or even a lifetime. It encompasses our knowledge, experiences, and skills, which are crucial in shaping our identities and capabilities. Over time, the three stages of information accessibility work together to enable the intricate process of human memory.

Another focal point in memory research centres on examining memory capacity and its limitations. Human memory has varying degrees of capacity across the three stages mentioned above. Sensory memory can briefly retain much sensory information (basically every signal it can sense) within its minimal duration. However, short-term memory further limits the amount of information retained. A widely accepted capacity limit for short-term memory is known as the “magic number seven” or “Miller’s law,” which suggests that individuals can hold approximately seven items, plus or minus two, in short-term memory (Miller, 1956). Also, it could be less to about four items if the retained information under operation (Cowan et al., 2005). In contrast, long-term memory is theoretically infinite in capacity (Atkinson & Shiffrin, 1968; Malmberg et al., 2019), although it is not without limitations. The challenge in long-term memory lies in retrieving information from forgetting it due to interference over time (Malmberg et al., 2019; Connerton, 2008). Memory researchers explore these capacities and limitations to understand better how the human mind processes information across different stages of memory.

Associative memory involves forming connections between pieces of information and maintaining them within the constraints of limited memory capacity (Raaijmakers & Shiffrin, 1981). This type of memory processes associations between concepts over short or long durations, playing a crucial role across all three memory stages.

In sensory memory, associative memory helps individuals link brief sensory inputs. For example, sensory memory is associated with auditory and visual stimuli when a bell-like sound is heard while a door opening is observed. Although both inputs are momentary, this process creates a transient connection, representing them as a unified sensory event (e.g., someone is coming in).

In short-term memory, related information can be grouped through a process known as chunking (Cowan et al., 2005; Mathy et al., 2018; Ma et al., 2023). For example, it allows us to reduce the number of memory units required to remember a telephone number by grouping seven individual digits into two units (e.g., 555-1234). This process temporarily maintains these associated number groups as integrated units for easy access.

In long-term memory, associative memory is vital in forming networks of related information. For example, the word “apple” might automatically trigger related concepts such as “fruit,” “red,” “sweet,” or “fall season,” forming a complex network stored in long-term memory. Another related example is language acquisition, whether for a first or second language, where speakers often form associations between words and their meanings or between words and their pronunciation. Language learners may also associate a new word in the target language with an image, a synonym, or specific personal experiences. Especially, vocabulary in a second language can be associated with the learner's native language. These associations help learners encode and retrieve words effectively during communication. In summary, associative memory is a crucial cognitive function that enables individuals to adapt to the diverse demands of their environment.

Learning is the initial stage of memory formation. New information is encoded into the intricate connection network among these learned items created by associative memory (Melton, 1963; Matzen et al., 2015; Klein, 2018). During learning, information is absorbed from stimuli and converted into a format suitable for later retrieval. This encoding process involves forming associations between newly learned items. These associations function as threads that connect related pieces of information, forming a cohesive network. This network of connections enables the effective recall of newly perceived information, enhancing individuals' ability to apply learned items across various contexts and adapt to different environments. In essence, associative memory plays a crucial role in encoding by weaving new information into individuals' existing network of learned items.

Learning performance may not rely solely on memory processes but often involves interaction with the environment, such as using tools for assistance. Fundamentally, the learning process entails encoding information from stimuli into mental records, facilitated by reorganising neural connections within specific nervous system regions. This process ensures that learned information remains accessible when needed. However, a contemporary perspective suggests that maintaining access to learned information does not rely solely on internal cognitive processes; it can also involve external elements in the physical world (Risko & Gilbert, 2016). For example, in the previous example, to keep the telephone number available, an individual might jot it down on paper, which serves the same function as remembering it for future use. When the cognitive process of learning is considered internal, a parallel physical action that serves the same purpose can be termed an "external process." Then, the learner's action to initiate this external process is termed "*externalisation*." This phenomenon is common in everyday life. In the previous example, choosing to jot down the telephone number instead of remembering it is an instance of externalisation.

A conceptual illustration comparing associative items in short-term memory, long-term memory, and physical media is shown in Figure 1. In short-term memory, only approximately three of the original five associative items can be retained due to capacity limitations. In long-term memory, associative items may not be retained correctly due to interference over time. In contrast, associative items externalised into physical media can be preserved relatively stably.

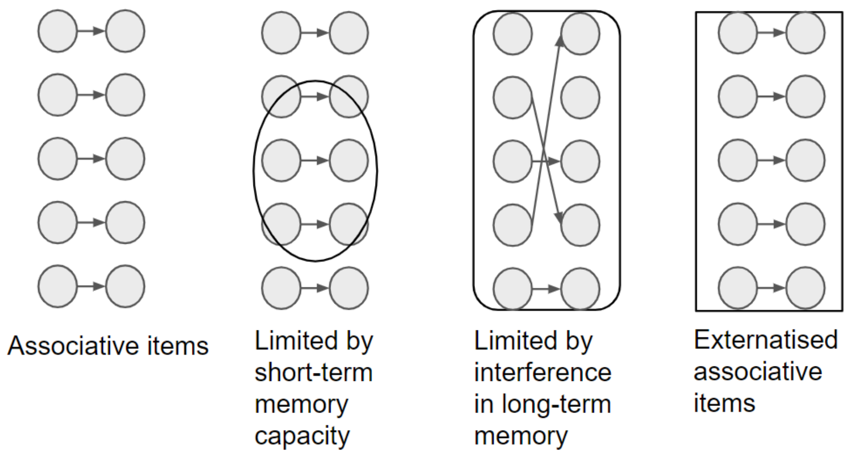


Figure 1. A conceptual illustration of associative items stored in short-term memory, long-term memory, and physical media. Each node represents an individual object, and two objects linked by an arrow represent an associative item. The circular area around the associative items represents the range of short-term memory. The rectangular area with rounded corners around the associative items represents the range of long-term memory. The rectangular area with sharp corners around the associative items represents the range of physical media.

Externalisation is a strategic trade-off in cognitive resource consumption rather than an unlimited, cost-free alternative to learning. It often involves transforming learned information from one format to another, typically choosing a format that is easier to store so that it reduces the mental workload required for memorisation. For example, jotting down a telephone number transforms the information from a numerical sequence to an episodic memory of where (or on which piece of paper) the numbers were written. As learning tasks become more complex, optimising memory regulation becomes increasingly essential for enhancing learning performance. Factors influencing individuals' choices to engage in externalisation are becoming a central focus of fundamental research on human memory, leading to the concept of *cognitive offloading*. This research trend has garnered increasing interest in recent years. This dissertation explores factors related to individuals' memory performance influencing their decision to employ cognitive offloading in learning tasks. Specifically, the current research defined short-term memory as the storage of items within 60 seconds, in line with initial findings by Prisko (1963). Accordingly, this work examined the relationships between long-term associative memory and the decision to engage in cognitive offloading. An overview of cognitive offloading and its determinants is provided in the following subsection.

1.2 Cognitive offloading

Many individuals use external resources to ease the cognitive burden of challenging memory tasks. These resources include media such as sticky notes or digital devices, which store information that would otherwise need to be remembered. In everyday life, individuals may prefer using a password manager to automatically access their accounts, writing a shopping list before going to the supermarket rather than relying on memory, or labelling group photos with names as a memory aid for future recall. This approach, known as cognitive offloading, has existed for thousands of years but has gained notable attention in recent years, primarily due to the widespread adoption of digital technologies such as smartphones, tablets, and navigation tools for various everyday tasks (Risko & Dunn, 2015; Risko & Gilbert, 2016; Grinschgl, 2020; Ma & Fujinami, 2024). As society increasingly adopts advanced cognitive offloading techniques through technology, this transformation not only reshapes daily routines but also creates opportunities for digital product development, underscoring the need for comprehensive research into cognitive offloading as a learning strategy.

One area of research examines the factors influencing individuals to employ cognitive offloading strategies instead of solely relying on mental effort (Risko & Dunn, 2015). This decision is guided by a cost-benefit analysis for each option to achieve a cognitive goal (Risko & Dunn, 2015; Dunn & Risko, 2016), considering external and internal determinants (Grinschgl, 2020).

External factors include aspects of the surrounding environment that either facilitate or hinder the selection of cognitive offloading. These factors play crucial roles in decision-making (Pereira et al., 2022). For example, the user-friendliness of external memory aids, such as note-taking apps or voice recorders, can significantly affect their adoption. Individuals may be more inclined to offload information when the technology is intuitive. Conversely, if tools are cumbersome, such as a digital planner that frequently crashes, has slow loading times, or is with a complex interface requiring multiple steps to access, individuals may choose mental effort instead.

The complexity of task demands also influences the decision to engage in cognitive offloading. Highly intricate tasks with intense concentration may prompt individuals to rely on external aids to prevent cognitive overload. Conversely, when a task is less complex, individuals may remember information directly, especially if doing so is more manageable than offloading it to external tools. For example, a person might use a digital calendar to track multiple upcoming events but rely on memory for a single appointment later in the day.

The level of environmental distraction is another external factor that can impact cognitive offloading (Gilbert, 2015a; Gilbert et al., 2023). Individuals may prefer to offload information in noisy or chaotic settings to avoid losing essential details. In quiet environments, they might rely more on memory and be confident in their ability to concentrate. For example, a student studying in a busy café might write down key points to avoid forgetting what they learned due to distractions. In contrast, in a quiet library, they might rely solely on memory, supported by a peaceful environment.

Moreover, culture and social norms influence whether individuals feel comfortable using external aids. In some cultures, seeking external assistance for memory-related tasks is widely accepted, whereas in others, it may be viewed as a sign of incompetence.

While these external factors significantly influence the selection of cognitive offloading, their direct contribution to optimising memory regulation may be limited, primarily because it is often challenging to alter them in practice. For example, an individual might have access only to an outdated computer, even though it is widely understood that a newer computer would be more suitable for completing the task. Consequently, individuals frequently adapt to existing conditions rather than control the environment based on personal preferences to select learning strategies. These external factors collectively shape the context in which cognitive offloading decisions are made, and understanding their impact is essential for comprehending how individuals navigate memory-related tasks across different environments. Furthermore, research on external factors provides insights that enhance the investigation of internal factors by clarifying the conditions influencing individuals' strategy choices under stable variables.

Internal factors include cognitive processes, particularly those related to memory functions, which are central to this dissertation. When external factors are stable, individuals consider their memory capabilities when deciding whether to employ cognitive offloading. For example, consider an employee in a professional setting tasked with remembering numerous client requirements and project deadlines. If employees have strong short-term memory and can mentally retain and manipulate this information, they may choose not to use external aids such as task management software. Conversely, another employee with less efficient short-term memory may have to use digital tools to offload details, thereby improving their project management effectiveness. In this way, the efficiency of individuals' memory function directly influences their decisions to employ cognitive offloading, allowing them to adapt their approaches based on cognitive functioning.

Importantly, decisions about cognitive offloading are influenced not only by actual memory performance but also by individuals' perceptions of their memory performance. In essence, individuals cannot directly access their actual memory performance; instead, they may have to rely on self-assessment of how well they process information. This self-assessment reflects perceived memory performance, which is expected to align with actual memory performance. The self-assessment significantly influences whether individuals rely on their cognitive abilities or use external aids to support memory and learning. For example, an individual working on a complex project may feel confident in their ability to remember key details, leading them to forgo task management tools. In contrast, another individual may doubt their capacity to recall all necessary information accurately, prompting them to rely more on digital tools as a memory aid. However, their actual memory performance might be similar if assessed through a memory exam, such as a task requiring recalling a list of paired associates. In this case, differences in strategy choices may stem directly from variations in perceived memory performance. Whether accurate or not, this self-assessment influences individuals' to optimise memory performance. It highlights a connection between cognitive offloading strategies and *metacognition*, which pertains to individuals' awareness and understanding of their cognitive processes. Risko and Gilbert (2016) proposed a model to describe the role of metacognition in deciding to employ cognitive offloading as a learning strategy (see Figure 2).

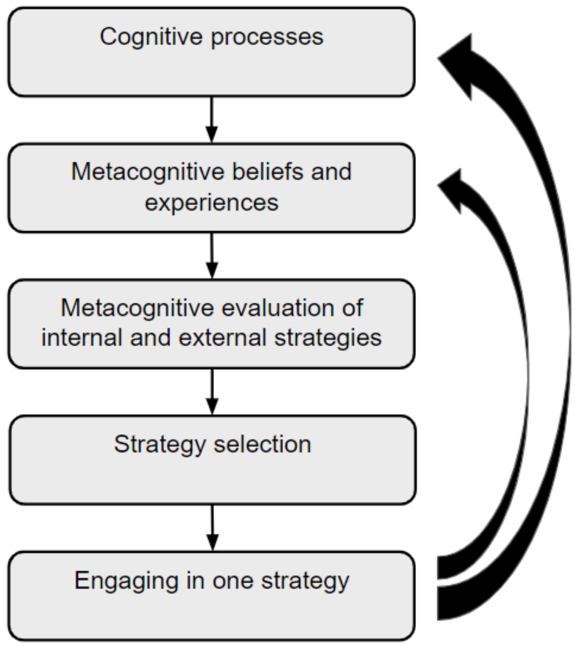


Figure 2. Risko and Gilbert's (2016) metacognitive model of cognitive offloading.

Metacognition represents a rich source of determinants influencing the choice to employ cognitive offloading (Flavell, 1979; Brown, 1978; Risko & Gilbert, 2016; Gilbert et al., 2023; Risko & Dunn, 2015; Dunn & Risko, 2016; Arango-Muñoz, 2013). The involvement of metacognition complicates memory performance by incorporating decision-making into the core memory processes of encoding, storage, and retrieval. This change prompts debate about the selection of cognitive offloading, specifically whether it is beneficial or detrimental in practical learning processes. Regarding this concern, research on the consequences of cognitive offloading has identified both positive and negative effects (Risko & Gilbert, 2016). For example, Grinschgl et al. (2021a) reported that employing cognitive offloading can enhance immediate performance on the Pattern Copy Task (Ballard et al., 1995) but may have detrimental effects on subsequent memory performance for the offloaded information. After all, attitudes toward cognitive offloading largely depend on the purpose of learning tasks, making it difficult to simply advocate for or against its use. Instead, a neutral stance may be more appropriate, although some research settings attempt to restrict cognitive offloading for specific research objectives (e.g., Gilbert et al., 2020; Ma & Fujinami, 2024).

This work examined the relationship between metacognition and strategy choices in memory tasks that involve cognitive offloading, aiming to explore how certain metacognitive factors interact to influence the selection of cognitive offloading. The following subsection provides an in-depth overview of metacognition and its relationship to cognitive offloading.

1.3 Metacognition, monitoring, and control

Metacognition, delineated by pioneering psychologists such as Flavell (1976, 1978, 1979) in developmental psychology, Brown (1978, 1980) in educational psychology, and Hart (1965, 1967) in cognitive psychology, pertains to individuals' awareness and comprehension of their cognitive processes. Generally, metacognition consists of three components, namely, *metacognitive knowledge*, *metacognitive experience*, and *metacognitive strategies*, which interact with two processes: *metacognitive monitoring* and *metacognitive control* (Flavell, 1979; Nelson & Narens, 1990, 1994; Norman et al., 2019).

Metacognitive knowledge represents individuals' understanding of their cognitive processes and how they function. This encompasses what they know about their thinking, learning, and problem-solving abilities. For example, metacognitive knowledge includes recognising the importance of practice in skill development or understanding the role of self-assessment in learning. On the other hand, metacognitive experience reflects individuals' real-time subjective awareness of their cognitive processes. It encompasses their feelings, judgements, and confidence levels regarding their cognitive abilities. Positive metacognitive experiences can motivate individuals to approach challenging tasks, whereas negative experiences may foster hesitation. Metacognitive strategies represent the techniques individuals use to enhance their cognitive performance. These strategies include setting goals, using memory aids or problem-solving tactics, etc. For example, an individual might apply a metacognitive strategy by breaking a complex project into smaller, manageable tasks to improve overall efficiency.

In addition to these three components, metacognition involves two dynamic processes: metacognitive monitoring and metacognitive control. Metacognitive monitoring refers to actively observing and evaluating cognitive processes during a task. For example, when solving a complex problem, an individual might notice that they are struggling to find a solution and, in turn, decide to rethink their strategy. On the other hand, metacognitive control involves actively regulating cognitive processes based on information gathered through monitoring. A theoretical framework describing these components and processes of metacognition was proposed by Nelson and Narens (1990) (see Figure 3). A practical example of this framework is that if someone realises (informed by metacognitive monitoring) that they are struggling to understand a concept while studying, they might look up additional resources or seek clarification from a teacher. In summary, metacognition comprises the interconnected components of metacognitive knowledge, metacognitive experience, and metacognitive strategies, all influencing the processes of metacognitive monitoring and metacognitive control.

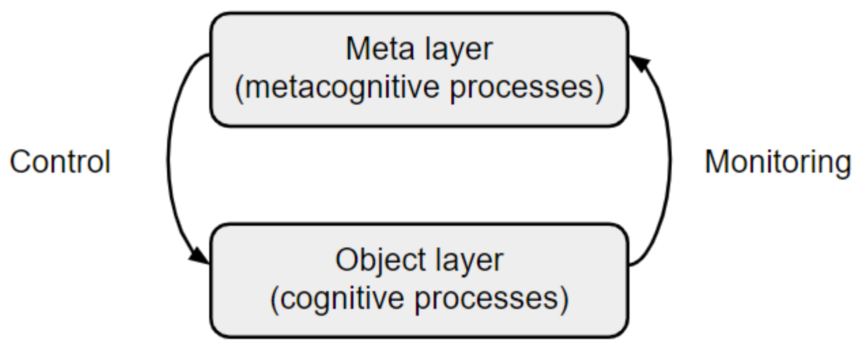


Figure 3. Nelson and Narens's (1990) theoretical framework of metacognition.

In cognitive psychology, the assessment of metacognition reflects the information processed during metacognitive monitoring through an *introspective* approach, including *retrospective* and *prospective* reports. Nelson and Narens (1990, 1994) and Nelson (1996) systematically described the role of subjective reports about introspection in their theory of monitoring and control. They outlined how retrospective and prospective reports are elicited across three stages of the memory process (i.e., encoding, storage, and retrieval): a retrospective report provides a judgement of a previous recall response, whereas a prospective report indicates a judgement of an anticipated recall response (Nelson & Narens, 1990). Introspective monitoring has become a significant methodological approach in cognitive psychology for examining metacognitive monitoring and control processes.

Metacognitive judgements refer to methods for eliciting retrospective and prospective reports, which have garnered significant research interest. These judgements assess individuals' perceived memory performance through questions posed during active encoding of to-be-remembered targets. Example questions might include "Do you feel you can remember this?" with a yes/no binary answer or "Do you feel confident in remembering this?" with a 0-100 numerical answer. Researchers in cognitive psychology expect responses to metacognitive judgements to correlate with individuals' actual memory performance and strategy choices (Arbuckle & Cuddy, 1969; Koriat, 1997; Rhodes, 2016; Norman et al., 2019). Specifically, studies view the connection between metacognitive judgements and actual memory performance as monitoring, whereas the connection between metacognitive judgements and strategy choices reflects control (Nelson & Narens, 1990).

The control process in research is often explored through learning behaviours based on two strategies: learning the target within a standard timeframe or adopting a more conservative approach. Examples of conservative strategies include requesting additional time or conducting more frequent reviews to enhance future learning performance. From this perspective, a relationship may exist between metacognitive judgements and the regulation of learning strategies. For example, optimistic metacognitive judgements tend to require less preparation time for memory tasks than pessimistic judgements do, demonstrating a reasonable monitoring-control relationship. Related research has reported similar findings (Dunlosky & Connor, 1997). This dissertation considers cognitive offloading as a conservative strategy within the monitoring-control relationship based on a negative relationship identified in previous literature (Son & Metcalfe, 2000; Metcalfe, 2009; Kornell & Bjork, 2007; Fiedler et al., 2019): "*individuals who feel that they cannot remember a target are more likely to offload the information than those who feel that they can remember it*" (Ma & Fujinami, 2024). Findings related to this relationship could have practical implications, as individuals often choose a learning strategy from multiple options based solely on their monitoring.

Ideally, metacognitive judgements should indicate actual memory performance with good monitoring *accuracy*, namely, the consistency between learners' perceived and actual memory performance should be high. For example, individuals with optimistic metacognitive judgements tend to exhibit better subsequent memory performance than those with pessimistic responses.

However, metacognitive judgements often lack accuracy. Related research commonly uses Goodman and Kruskal's gamma correlation coefficient to measure the accuracy between perceived and actual memory performance (Rhodes, 2016), typically finding values below 0.4 (Nelson & Dunlosky, 1991; Tauber & Rhodes, 2012; Hu et al., 2016). Distorted metacognitive judgements can lead to *overestimation*, where individuals believe that they have learned material more effectively than they have, resulting in perceived memory performance that is higher than actual memory performance. Conversely, metacognitive judgements can also exhibit *underestimation*, where individuals report perceived memory performance more conservatively than actual memory performance. Discrepancies between perceived and actual memory performance can have unexpected consequences. Overestimation may lead to complacency, where individuals neglect effective learning strategies because they believe they have already mastered the material. Conversely, underestimation might result in unnecessary, time-consuming reviews of material they have already learned well. Both types of inaccuracy may cause learners to use less effective strategies in managing their time and resources.

This issue is particularly evident in the context of cognitive offloading. Distorted metacognitive judgements can obscure the likelihood of choosing a conservative learning strategy, making the selection of cognitive offloading unpredictable. Several studies reflect specific aspects of this problem. For example, the relationship between metacognitive judgements and cognitive offloading is sometimes weak (Grinschgl, 2020; Sachdeva, 2023), lacks a causal basis (Grinschgl, 2020; Grinschgl et al., 2021b), and is unresponsive to improved metacognitive accuracy through metacognitive intervention (Engeler & Gilbert, 2020; Gilbert et al., 2023). Inaccurate metacognitive judgements may hinder the monitoring process, reducing the ability to predict the selection of cognitive offloading in learning scenarios and, consequently, limiting potential learning gains that individuals could otherwise acquire. Additionally, the practical effectiveness of cognitive offloading as a learning strategy may be compromised when monitoring accuracy is limited. Therefore, improving monitoring accuracy is essential.

1.4 First- and second-order metacognitive judgements

Individuals can reassess their previous metacognitive judgements by questioning the accuracy of their perceived memory performance, using the answer to evaluate the quality of their initial judgement. From this perspective, the initial assessment of perceived memory performance is a *first-order metacognitive judgement* (FOJ), and the reassessment is called *second-order metacognitive judgement* (SOJ). A common method for eliciting an SOJ is to ask individuals to rate their confidence in the correctness of their previous FOJ responses through prompts such as "Please indicate your confidence in your last performance estimation on the following scale." The concept of "metacognitive judgement" thus includes both the FOJ and the SOJ.

Dunlosky et al. (2005) initially proposed the concept of the FOJ and the SOJ from their *dual perspective*, suggesting that combining FOJs and SOJs may improve monitoring accuracy. This dual perspective outlines a two-phase monitoring process in learning tasks, namely, individuals first report their perceived memory performance as the FOJ, followed by a confidence rating of the FOJ as the SOJ to assess its accuracy (Dunlosky et al., 2005). Furthermore, the SOJ may also regulate FOJ within the same learning task (Buratti & Allwood, 2012; Buratti & Allwood, 2015). Previous work has considered FOJ as perceived performance at the object (i.e., actual memory) layer. In contrast, the SOJ is regarded as perceived performance at the meta (i.e., perceived memory) layer. Similarly, other studies also consider the SOJ as a higher meta layer (Nelson & Narens, 1990) or a meta-meta layer (Buratti & Allwood, 2015), in contrast to the FOJ at the object layer. An example of the FOJ-SOJ relationship is given in previous work (Ma & Fujinami, 2024): "*when individuals in a learning task judge whether they can recall a specific item, this assessment reflects their actual memory performance and is considered an FOJ. Subsequently, when they evaluate how confident they are in the accuracy of that assessment, they make an SOJ, which is a reassessment based on their earlier assessment of actual memory performance (i.e., perceived memory performance).*" The FOJ-SOJ relationship reflects the role of metacognition in regulating cognition (Nelson & Narens, 1990, 1994; Dunlosky et al., 2005), as shown in Figure

4. Dunlosky et al. (2005) explained why the SOJ can reflect FOJ accuracy, suggesting that forming an SOJ accumulates evidence to support previous perceived memory performance, as indicated by the magnitude of confidence levels reported on the SOJ scale. SOJs hold promise for evaluating monitoring accuracy, as the process provides individuals with additional chances to assess the consistency between their perceived and actual memory performance.

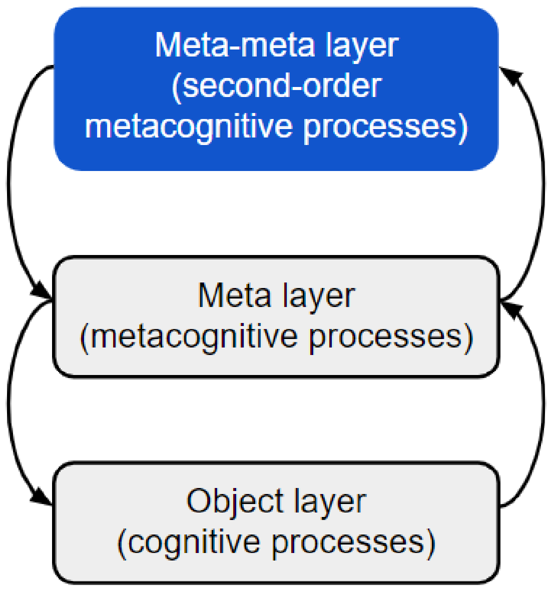


Figure 4. The concept of higher-order metacognition, assessed by SOJs. This new layer of metacognition is highlighted in blue.

The FOJ-SOJ relationship is supported by empirical evidence. In the related literature, Dunlosky et al. (2005) reported that this relationship follows U-shaped curves when both perceived memory performance and confidence level are elicited by a 0-100 numerical scale. Specifically, SOJ magnitudes peak at the two extreme values of FOJs. This nonmonotonic tendency between perceived memory performance and confidence level suggests that FOJ and SOJ formations operate independently. A recent study also provides evidence for nested metacognitive judgements up to the third order (i.e., meta-meta-meta layer). It explains this structure through a computational model (Recht et al., 2022). SOJs can be considered theoretically and empirically valid for monitoring FOJs.

SOJs have been widely incorporated to increase monitoring accuracy across various psychological disciplines, including educational psychology (Miller & Geraci, 2011; Händel & Fritzsche, 2016; Fritzsche et al., 2018), developmental psychology (Serra et al., 2008; Nederhand et al., 2021), and personality psychology (Buratti et al., 2013). For example, research in educational psychology has shown that SOJ reports can detect overestimated FOJs in students with relatively low academic performance (Miller & Geraci, 2011; Händel & Fritzsche, 2016; Fritzsche et al., 2018). In developmental psychology, the ability to correct FOJs via SOJs is absent in secondary school students (Nederhand et al., 2021) and varies significantly between younger and older adults (Serra & Dunlosky, 2008). Additionally, research on personality psychology suggests that the relationship between FOJs and SOJs differs among individuals with varying personality traits and cognitive styles (Buratti et al., 2013). These studies revealed that confidence levels interact with the relationship between perceived and actual memory performance.

It is reasonable to assume that confidence level might similarly interact with perceived memory performance in the selection of cognitive offloading. The relationship between FOJs and the selection of cognitive offloading has been reported consistently throughout the literature (Risko & Dunn, 2015; Gilbert, 2015b; Boldt & Gilbert, 2019; Hu et al., 2019; Scott & Gilbert, 2024). On another note, while incorporating SOJs can improve monitoring accuracy by correcting distorted FOJs, the role of SOJs in addressing FOJ inaccuracies when deciding to employ cognitive offloading as a learning strategy remains unexplored. This dissertation was motivated by this gap, aiming to investigate the influence of SOJs on the relationship between FOJs and strategy choices, particularly when cognitive offloading is an option. A summary of key literature in related research can be found in Table 1.

Table 1. Summary of key literature in related research.

Historical development	Literature
First, conception of metacognition	Hart (1965, 1967); Flavell (1976, 1978, 1979); Brown (1978, 1980)
Afterwards, theoretical framework: monitoring-control architecture	Nelson & Narens (1990, 1994); Nelson (1996)

Meanwhile, empirical monitoring-control relationship	Dunlosky & Connor (1997); Son & Metcalfe (2000); Metcalfe (2009); Kornell & Bjork (2007); Fiedler et al. (2019)
On the other hand, inaccurate monitoring (inconsistent performance was found) (Problem)	Nelson & Dunlosky (1991); Tauber & Rhodes (2012); Hu et al. (2016); Rhodes (2016)
Then, reassessment for improved monitoring accuracy was identified (Approach)	Dunlosky et al. (2005); Serra et al. (2008); Miller & Geraci (2011); Buratti & Allwood (2012, 2015); Buratti et al. (2013); Händel & Fritzsche (2016); Fritzsche et al. (2018); Nederhand et al. (2021)
Finally, reassessment for new monitoring-control relationship was motivated (Originality)	Ma & Fujinami (2024)

1.5 The present study

This work raises this research question by recognising that the monitoring process in a learning task involves both perceived memory performance and corresponding confidence levels. Specifically, this dissertation examines how individuals decide to employ cognitive offloading based on two layers of self-assessment. The question was investigated via paired-associate learning tasks, where participants learned English word pairs (e.g., “ABILITY” is the cue and “CAPABILITY” is the target for the associate “ABILITY – CAPABILITY”), following a paradigm similar to that employed in (Dunlosky et al., 2005)’s work. An essential adaptation in this work’s paradigm was allowing participants to employ cognitive offloading during the Learning session and access offloaded information during the Test session. Two hypotheses were formulated accordingly.

First, a negative correlation between FOJs and the selection of cognitive offloading was anticipated, in line with findings from prior research (Risko & Dunn, 2015; Gilbert, 2015b; Boldt & Gilbert, 2019; Hu et al., 2009). FOJs were elicited with questions such as “Do you feel you can remember it?” (For details, see Chapter 2 Methodology). The responses were limited to “YES” or “NO,” as in related studies (Jersakova et al., 2017; Zheng et al., 2023) because this binary format integrates more efficiently with multiple levels of SOJs than higher-resolution FOJ scales do. Hypothesis 1 proposes that individuals are likelier to offload a paired associate on trials where they respond “NO” to the FOJ than on trials where they respond “YES.”

Second, an interaction effect of SOJs on the strength of the association between FOJs and strategy choices was anticipated. An SOJ was elicited immediately following the FOJ on the same trial via a 0-100 numerical scale, with questions such as "Please report how confident you are in the last 'YES or NO' answer?" (for details, see Chapter 2 Methodology). This reassessment should capture the monitoring process at a higher order, similar to the approach employed by Dunlosky et al. (2005), which illustrates the two-phase monitoring process. Hypothesis 2 proposes that the negative association between FOJs and strategy choice is weaker in trials with lower SOJs than in trials with higher SOJs. A concept flow for the two hypotheses is shown in Figure 5.

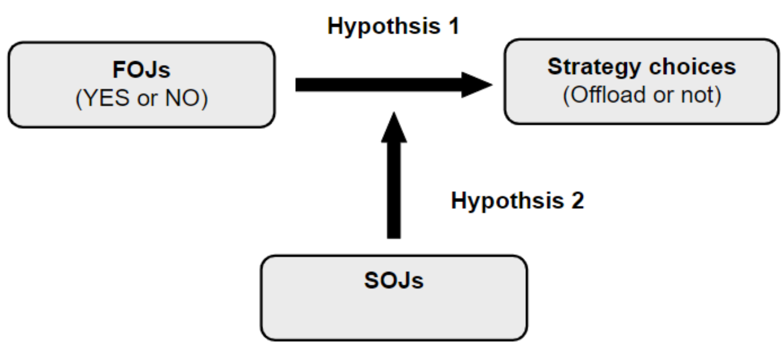


Figure 5. A concept flow for the illustration of two hypotheses.

The remainder of this dissertation is organised as follows: Chapter 2 outlines the methodology; Chapter 3 presents the results of two pilot studies (Studies 1 and 2); Chapter 4 presents the results of the main study (Study 3); Chapter 5 interprets the findings and discusses potential issues; and Chapter 6 provides the conclusion, outlining contributions, impacts, implications, and suggestions for future research.

A specific structure of this dissertation is outlined by a flow chart shown in Figure 6. Chapter 2 presents predetermined parameters, including the duration of each learning task, the number and difficulty of paired associates, and other relevant criteria. Expectations regarding the conditions necessary to ensure the validity of these parameters are also mentioned, along with one additional question concerning detailed research design. Chapter 3 then introduces two pilot studies: Study 1, which validates the predetermined parameters, and Study 2, which answers the additional research question. Furthermore, the second experimental condition of Study 2 (i.e., Study 2B) was selected for continuation in the main study. This condition was applied to a larger sample size as Study 3, which is reported in detail in Chapter 4, interpreted in Chapter 5, and summarised in Chapter 6.

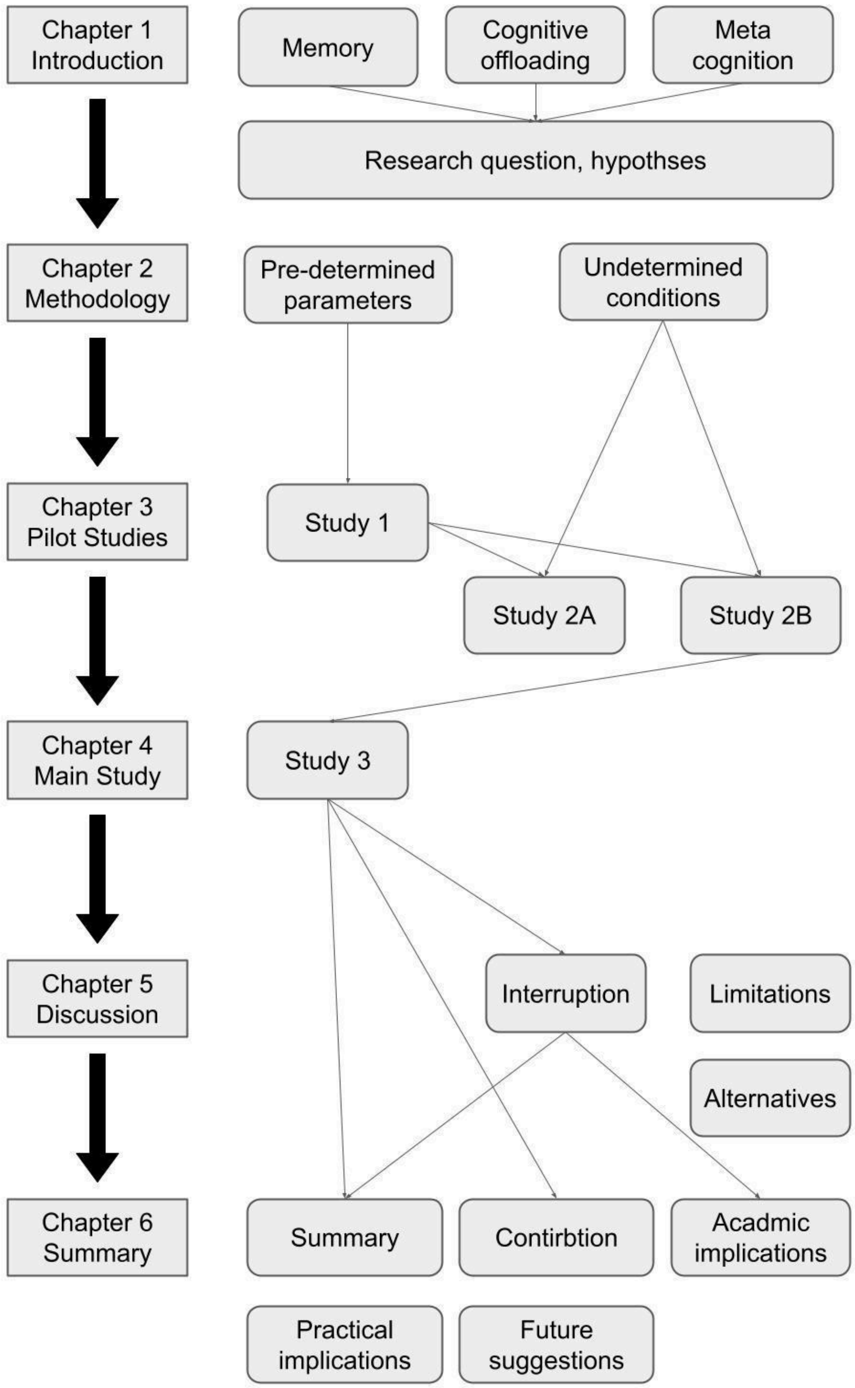


Figure 6. The flowchart illustrates the dissertation structure. Rectangles with sharp corners represent chapters, whereas rectangles with rounded corners represent components within each chapter. Bold arrows show the progression from the first to the last chapter, and thin arrows show the connections between chapter components.

Chapter 2 Methodology

This dissertation employed a variant of the paired-associate paradigm to investigate the relationships between individuals' metacognitive judgements and their selection of cognitive offloading. In this paradigm, participants were instructed to memorise paired associates, each consisting of a cue and a target (e.g., ABILITY – CAPABILITY), for later recall on an item-by-item basis. It can be divided into three sessions – Learning, Retention, and Test. This paradigm is a standard method for exploring various aspects of human memory and learning processes (Richardson-Klavehn & Bjork, 1988; Van Woudenberg, 2008). Thorndike (1908) initially used this paradigm to investigate the memory of association between two objects (i.e., associative memory). Hart (1965) also used this paradigm to explore the consistency between metacognitive judgements and monitoring accuracy. In recent years, research has also employed it in the relationship between metacognition and the selection of cognitive offloading (Hu et al., 2019). In the current work, the paradigm was developed in a program using an online experiment builder, Gorilla, and implemented by recruiting participants remotely from the United States via another online platform, Amazon Mechanical Turk (MTurk). This program enables data collection related to first-order metacognitive judgements (FOJs), second-order metacognitive judgements (SOJs), recall responses for paired associates, and the decision to employ cognitive offloading. Further details are provided in the following subsections.

2.1 A Learning session

The Learning session involves presenting participants with pairs of English words and instructing them to memorise the association between the two words. For example, participants may be shown the paired words "ABILITY – CAPABILITY" and asked to remember it as a paired associate for the later test. Various factors may influence the design of this learning paradigm.

The difficulty level of paired associates for memorisation is typically a crucial factor, which is primarily determined by the strength of the semantic association between two words in both *forward* and *backward* directions. The strength of the association is based on the observed probability of two words appearing together, with the natural sequence being the forward association. In daily communication, we naturally encounter words in specific sequences, and this forward association aligns with our everyday linguistic experiences. In contrast, the reverse order, or backward association, is often less common. While forward association is generally more robust and vital, serving as the cornerstone for building meaningful connections and effectively utilising vocabulary, backward association also has secondary significance. The example associated "CAB – TAXI (strength for forward is 0.54 and for backward is 0.01, Nelson et al. (1998))" illustrates this concept: in the forward association ("CAB" to "TAXI"), "CAB" naturally leads to "TAXI," following common usage patterns and making it easier to remember. Conversely, the backward association ("TAXI" to "CAB") is less direct and less frequently encountered, making it harder to recall. The forward association enhances memory by aligning with natural language sequences, whereas the backward association provides secondary support. These influences are related to the difficulty of pairing, which must be carefully controlled.

Another consideration in paired-associate studies is the workload, which raises two key questions: (1) the time allocated for memorising each paired associate and (2) the optimal number of paired associates to include in a single session. Addressing these questions is critical to designing effective learning sessions. First, the duration allocated for memorising each paired associate directly influences the depth of memory consolidation. If the time is too brief, participants may be unable to form strong associations; if it is too long, it may lead to fatigue and reduced attention. Finding the right balance is essential to facilitate productive learning. Second, determining the optimal number of paired associates is crucial for managing the cognitive load. Including too many pairs in a single session can overwhelm participants, impeding their focus and reducing their ability to memorise the associations effectively. On the other hand, too few paired associates may limit the scope of learning. Thus, determining an ideal number of associates to include ensures a meaningful learning session without feeling overwhelmed or understimulated. These considerations are fundamental to structuring an effective learning session that optimally balances cognitive demands.

This study utilised paired-associate tasks in which participants were presented with uppercase English word pairs, with the left word serving as the cue and the right word as the target (e.g., "ABILITY" as the cue and "CAPABILITY" as the target for the pair "ABILITY – CAPABILITY"). Each participant was instructed to complete each paired-associate task within one minute. To ensure that the experiment lasted approximately one hour, 53 pairs were selected from the South Florida Free Association norms (Nelson et al. 1998; D. Nelson et al. 2004). According to the original norms, all selected associates exhibited weak semantic relatedness in the forward (0.01) and backward (0) directions, as the relative weak relatedness could avoid the potential reactivity of eliciting metacognitive judgements (Double et al., 2018). The forward strength was calculated by the percentage of participants responding to the target as the first word came to mind when the cue was given (i.e., free association). Similarly, the backward strength was calculated by the percentage of participants responding to the cue as the first word came to mind when the target was given. An example for the calculation of strength for "ABILITY – CAPABILITY" is shown in Table 2. The participants had five seconds to learn each pair of associates. Five of the 53 pairs were designated for practice, whereas the remaining pairs were used for formal study.

Table 2. The calculation of forward and backward strength for "ABILITY – CAPABILITY."

	Forward	Backward
Cue	ABILITY	
Target	CAPABILITY	
Question	ABILITY _________	CAPABILITY _________
Number of participants	143	124
Number of responses	17	35
Strength	0.12	0.28

The example for forward strength is summarised: given ABILITY _________, 17 out of the 143 wrote CAPABILITY on the blank next to it, obtaining 0.12 forward strength. The example for backward strength is summarised: given CAPABILITY _________, 35 out of the 124 wrote ABILITY on the blank next to it, obtaining 0.28 backward strength.

This example is from Nelson et al.'s (1998) descriptions provided through http://w3.usf.edu/FreeAssociation/Intro.html.

Even though all paired associates were selected based on the same forward and backward strength, several factors may still contribute to variations in learning difficulty across trials. These factors include word characteristics, word length, and phonological similarity, which may cause some pairs to be more intuitively linked than others. To control for these differences, trial order was randomised across participants in the Learning session.

2.2 A Test session

In each trial of the Test session, participants were instructed to answer the target of the paired-associate according to the demonstration of its cue. For example, participants may be shown an isolated association “ABILITY” of the paired-associate “ABILITY – CAPABILITY,” and they will then respond “CAPABILITY” as the correct answer of that trial.

Typically, two approaches for reporting the targets during the Test session are *recognition* and *free recall*. In recognition, participants are given the cue and asked to identify the correct target word from a list of options. For example, participants might be given the cue “ABILITY” and then presented with a list of word options, including “CAPABILITY,” “SKILL,” “TALENT,” and “POTENTIAL.” The participant must recognise and select “CAPABILITY” as the correct answer. In contrast, free recall requires participants to remember and produce the target word without any options. In this scenario, they might be given the cue “ABILITY” and asked to provide the associated target word without any prompts. The participants then responded with “CAPABILITY” from their memory, demonstrating their ability to recall the target word independently. These two reporting approaches, recognition and free recall, provide different perspectives on participants’ memory and understanding of paired associates, allowing researchers to assess memory retention through distinct cognitive processes.

Time limitations might bring extra pressure to participants, causing their learning performance to be influenced positively or negatively. For example, in a time-constrained testing environment, participants might experience positive pressure, leading to heightened focus in retrieving paired associates. This can result in quicker and more accurate responses when participants recall or recognise target words, demonstrating improved memory retention within a constrained time frame. Conversely, time limitations can also have adverse effects on testing performance. Participants might feel rushed, leading to a decreased ability to recognise or recall paired associates accurately. In such cases, the pressure of the constrained time frame can negatively impact participants' memory retrieval, potentially resulting in lower testing scores. The influence of time limitations in the Test session may vary among individuals and can be influenced by factors such as cognitive processing speed and stress management. Therefore, it is essential to consider the introduction of a time limitation in the Test session, as it can significantly impact the reliability of assessments and the overall testing experience.

In accordance with the above considerations, this work utilised pair-associate tasks with a Test session in which participants input the targets in free recall without time limitations. This approach gave participants a less pressured and more comfortable testing environment. In this Test session, the participants were encouraged to freely recall the target words associated with the given cues without the added stress of a constrained time frame. This method aimed to assess the depth of participants' memory retention by allowing them to recall paired associates in a manner that mirrored real-world memory use, remaining a relatively high challenge for memorisation. By eliminating time limitations, the focus shifted to the accuracy of memory recall, providing insights into the effectiveness of the Learning session.

When strategy choices are involved in the Learning session, the scoring methods for the Test session may have multiple reasonable approaches. The simplest method is to treat correct recall by both strategies as equal. Alternatively, correct recall based on offloaded information could be considered less valuable than recall from memory, given that employing cognitive offloading intuitively feels easier than relying on memory. Since the cognitive offloading strategy may be overused (Dunn & Risko, 2016), it could create an imbalance in the sample (in an extreme case, no trials may be completed using memory). Therefore, it is necessary to control the unnecessary use of cognitive offloading. Balancing scores across different learning strategies could reduce the tendency to employ cognitive offloading.

2.3 A Retention session

A Retention session in the paired-associate paradigm is needed between the Learning and Test sessions. The purpose of this session is to temporarily disconnect the learning and recall processes to ensure that the information retrieved during the Test session is from individuals' long-term memory. In this way, the memorised information can be recognised as practically valid.

The critical point for the Retention session is to add irrelevant cognitive tasks as one additional session between the Learning and Test sessions. As discussed in the previous paragraph, a Retention session in the paired-associate paradigm serves a crucial purpose in the learning process. It acts as a temporary disconnect between the learning and recall processes to ensure that the information retrieved during the Test session is from individuals' long-term memory, validating its practicality. This disconnection is made possible through the inclusion of irrelevant cognitive tasks. Introducing these tasks in the Retention session diversifies the participants' attention from the recently learned paired associates. This diversion helps prevent immediate recall and forces the participants to rely on their long-term memory, making the Test session a more accurate measure of memory retention. Therefore, the inclusion of irrelevant cognitive tasks plays a vital role in enhancing the validity of the entire paired-associate paradigm by ensuring that the information recalled during the Test session reflects the participants' genuine ability to retain and retrieve the learned associations.

The current work used simple arithmetic operations as distractor tasks to prevent participants from actively rehearsing paired associates during the retention interval. Specifically, participants were presented with 144 addition and multiplication calculations, each involving two randomly selected numbers from 1-9. These calculations were strategically designed to engage participants' cognitive resources in a different mental domain, ensuring that their attention was diverted from the recently learned paired associates.

The inclusion of distracting tasks serves the primary purpose of maintaining participants' optimal performance during the Learning and Test sessions. Since the distracting tasks only ensured that the participants did not actively rehearse the paired associates during the retention interval, these results were not analysed due to the irrelevance of the current work's objectives. The primary goal is to maintain the integrity of the learning and testing processes, allowing for a more accurate assessment of memory retention.

2.4 Scale types of metacognitive judgements and regulation

In this study, each task in the Learning session is followed by an FOJ, an SOJ, and a selection of learning strategies (employing cognitive offloading or not). After memorising a pair of associates, participants responded to an FOJ by answering, "Do you feel you can remember the paired associate correctly for the latter test?" Then, they responded to an SOJ by answering the following question: "Please indicate your confidence level in the last answer." Finally, they responded to the selection of learning strategies by answering the following question: "Please select your strategy between 'Writing down' and 'Keeping in Mind':" to determine whether they offload. The three questions are asked continuously.

Two main scale types correspond to an FOJ question: discrete in binary (yes/no) and continuous in percentage (0 to 100%). The choice between these two scale types depends on the specific objectives of the FOJ task. Binary scales, such as yes/no, offer a simplified and dichotomous approach, allowing participants to make straightforward decisions about their perceived learning. In contrast, continuous percentage scales provide a more nuanced means of expressing confidence in their learning. The participants could indicate their judgement on a sliding scale, reflecting a more comprehensive range of perceived learning, from absolute uncertainty (0%) to complete confidence (100%). The type of scale used can significantly impact the depth of insight gained from FOJ responses in a given study.

The consideration regarding the selection of the SOJ scale is similar: two scale types corresponding to an SOJ question – discrete in a few options (i.e., "Absolutely certain/Fairly sure/Possibly correct/Just a guess" for a four-option case (Crawford & Stankov, 1996) and contentious in percentage (0 to 100%). The choice of SOJ scale type hinges on the assessment's specific objectives and the desired precision level in capturing the participants' confidence level in their perceived memory performance. Four-point verbal scales provide a structured framework for participants to express their confidence or uncertainty in their responses. Conversely, continuous percentage scales offer a more granular approach, allowing participants to indicate their confidence level with a broader range of precision. This scale type is valuable when researchers require a more detailed understanding of participants' subjective states.

The selection of cognitive offloading is addressed through a binary option, providing participants with either "Writing down" or "Keeping in Mind." When participants opt for "Writing down" during the Learning session, they can leave a reference on a piece of paper they prepared themselves for use in the subsequent Test session. This option enables participants to strategically externalise part of the cognitive load, reducing the need for extensive memorisation while keeping vital information readily accessible for the recall task. In contrast, those who decide "Keeping in Mind" rely solely on their internal memory resources during the Learning session, which may require greater cognitive effort for retention and recall. This binary choice provides insight into individual preferences for cognitive offloading strategies and significantly shapes the processes involving learning and recall.

This work utilised a binary scale for FOJs and the selection of cognitive offloading to keep the data consistent. On the other hand, this work utilised a numerical scale for SOJs, as using different scale formats in two subsequent elicitations can help reduce potential bias (Mochon & Frederick, 2013). Additionally, different patterns in the selection of cognitive offloading might emerge if strategy choices are scored differently. For example, individuals might be more inclined to employ cognitive offloading if the strategies "Writing down" and "Keeping in Mind" were scored as equivalent, compared to scenarios where "Keeping in Mind" received a higher score than "Writing down." Chapter 3 explores this issue in detail (see Study 2).

2.5 Remote implementations and online platforms

The implementation of online experiments in this work relies on two main platforms – Gorilla (Figure 7) and MTurk (Figure 8) (Anwyl-Irvine et al., 2020; Keith et al., 2017). First, the program for the learning tasks was developed on Gorilla, an online experiment builder that provides a link for implementation. This link allows participants to run the program remotely via browsers on their desktop or laptop computers. The current work restricted mobile devices (such as smartphones and tablets) to ensure that participants had a relatively consistent viewing experience and were not disadvantaged by screen size or other technical limitations. The study materials were designed to adapt automatically to the size of the participants' screens to ensure they were easily readable. After the settings were determined, the link was shared on MTurk to hire participants with rewards.

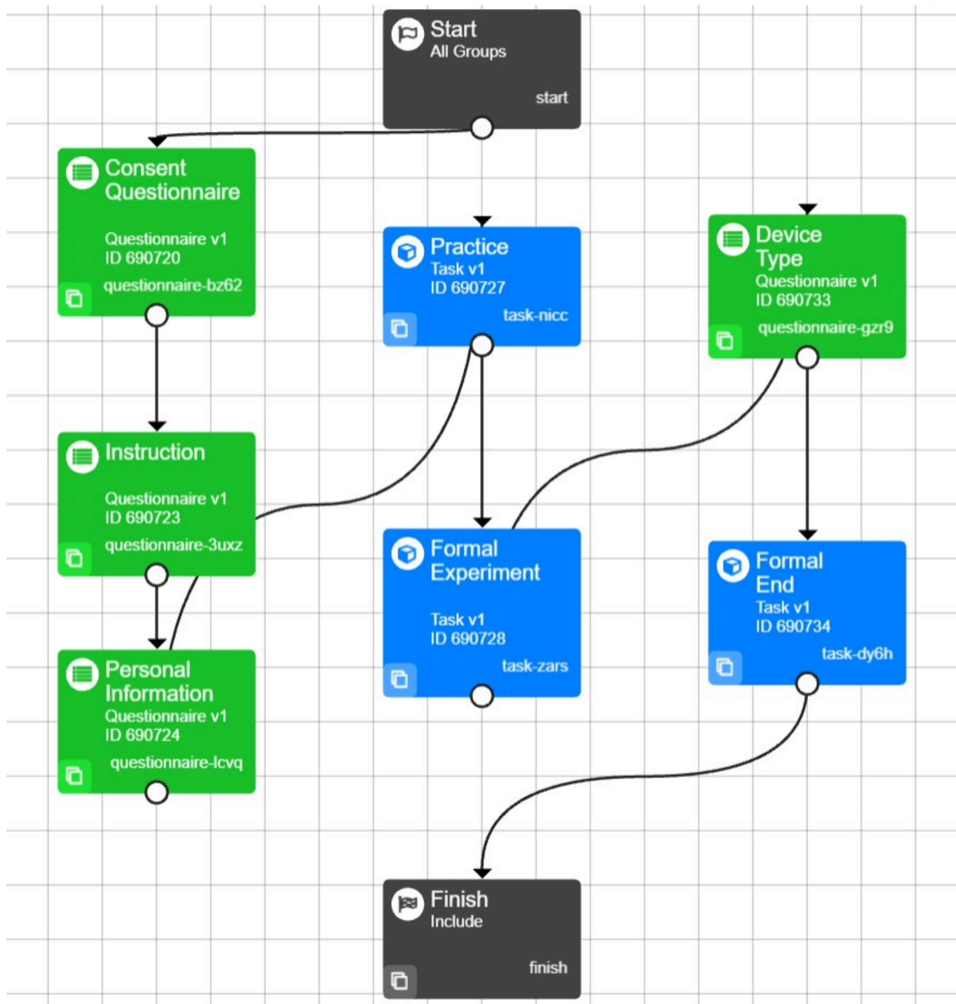


Figure 7A. An example experiment designed in the configuration screen on Gorilla.

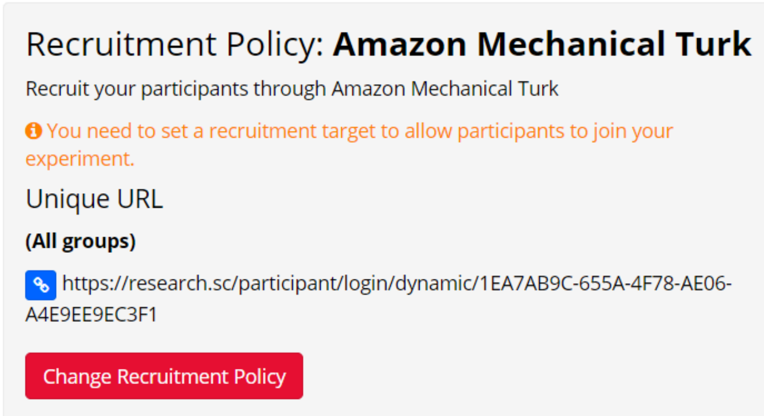


Figure 7B. The link generated by Gorilla for recruiting participants on MTurk.

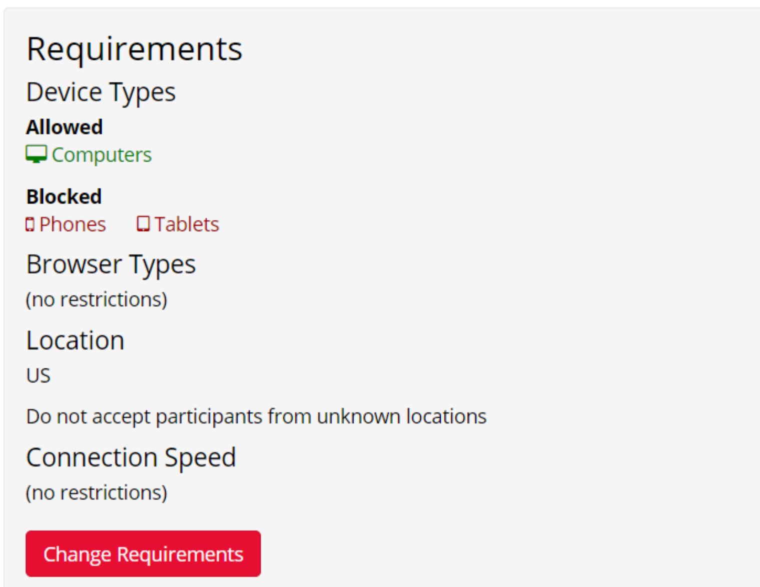


Figure 7C. The device limitation setting on Gorilla.

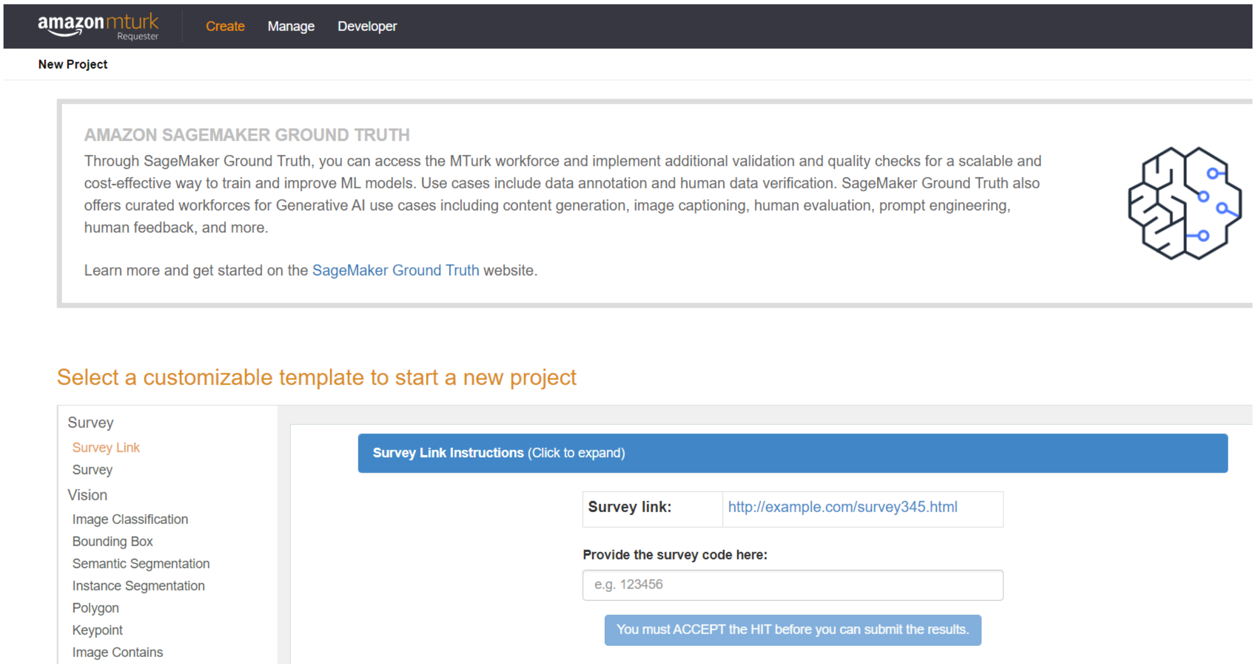


Figure 8. The area in which the remote experiment created by Gorilla on MTurk was distributed.

2.6 Participants

Individuals in different age groups may exhibit distinct performance patterns in learning tasks. For example, younger individuals may have more robust memory functions and faster learning capacities, allowing them to recall information more efficiently. In contrast, older individuals may have accumulated more life experiences, which can provide them with various strategies for learning tasks. However, they may face challenges related to age-related cognitive decline, affecting their ability to memorise information efficiently.

Individuals in various age groups may also have unique patterns of metacognitive judgements (Hines et al., 2009; Sanders & Berry, 2021). For example, children with limited life experience may exhibit less accurate metacognitive assessments, often overestimating or underestimating their learning capabilities. In contrast, adults tend to develop more refined metacognitive skills, allowing them to make more accurate judgements about their cognitive performance, leading to better self-regulation in learning and memory tasks. However, while older adults have accumulated a wealth of experiences, they may sometimes demonstrate alterations in metacognition due to age-related cognitive changes, potentially leading to misjudgements about their memory abilities.

Language ability is another unignored aspect of memorising paired associates, with significant differences observed among individuals with varying levels of familiarity with the language of those paired associates. Those highly proficient in a language have an advantage in memorisation. Their deep understanding of the language's grammar, vocabulary, and cultural connotations allows for more meaningful associations between paired associates. In contrast, individuals with limited language proficiency in the target language may struggle with memorisation, especially if they lack essential vocabulary. These language barriers can hinder the formation of solid associations and thereby impact memorisation efficiency.

The current work recruited participants aged 20-50 from the United States through MTurk. Consequently, all materials used in the study, as presented in this dissertation, are in American English. This work included only MTurk workers with a high reputation (HIT Approval Rate for all requesters' HITs greater than 99%) (Peer et al., 2014; Peer et al., 2022). The average participation time was within one hour, and each participant was compensated with USD 6 for their time.

2.7 Summary of the present research

This research included two pilot studies (Studies 1 and 2) and one main study (Study 3). Two pilot studies met specific expectations and confirmed that certain settings were appropriate. The predetermined parameters for the Learning session of paired associates included a learning period of 5 seconds, a forward strength of 0.1, and a backward strength of 0.0. For the Retention session, the parameters included 144 self-paced arithmetic tasks, and for the Test session, self-paced free recall was used. Additionally, the workload consisted of 53 paired associates, and would be completed by participants aged 20-50 years from the same English-speaking country. Moreover, the scoring methods for assessing cognitive offloading were compared. For details on the two pilot studies, see Chapter 3. On another note, the main study was conducted using parameters established in the two pilot studies to address the research question (Chapter 4). The main hypotheses across the three studies were verified using the paired-associate paradigm, along with the implications they reveal, as summarised in Table 3.

Table 3. Summarisation of hypotheses across three studies.

Study	Hypothesis	Implication
Study 1	1. Actual memory performance, measured by free recall in the Test session, was expected to result in less than half of the trials being recalled when using the predetermined parameters (a 5-second learning period, a 48-trial learning workload, a forward strength of 0.1, and a backward strength of 0.0) in the Learning session after the Retention session involving 144 arithmetic tasks.	Reflect that there is room for improvement in learning if cognitive offloading is employed
	2. Perceived memory performance, as measured by FOJs, was expected to differ significantly from actual memory performance, either being significantly higher or lower.	Reflect that FOJs are biased and highlight the need to address inaccuracies in metacognitive monitoring through SOJs
Study 2	3. The selection of cognitive offloading when using a balanced scoring method between two learning strategies for correct recall in the Test session was expected to be lower than when both learning strategies were treated as equal for correct recall.	Recommend the optimal choice for the main study that minimises unnecessary cognitive offloading
Study 3	4. A significant main effect of FOJs on the selection of cognitive offloading was expected.	Replicate previous findings suggesting that the monitoring process is inaccurate
	5. A significant interaction effect between FOJs and SOJs on the selection of cognitive offloading was expected.	Suggest that SOJs address inaccuracies in the monitoring process

Chapter 3 Pilot Studies

Overview

This dissertation aimed at the relationship between metacognition and the selection of cognitive offloading. Specifically, metacognition is regarded as the perceived memory performance measured by first-order metacognitive judgements (FOJs) and the confidence level measured by second-order metacognitive judgements (SOJs). In this sense, the challenge of learning tasks is crucial. They should be challenging enough to require cognitive effort but not so difficult that they discourage memory use. Ideally, tasks should be at a moderate difficulty – neither too easy, making offloading feel redundant, nor too hard, leading participants to rely entirely on offloading. Therefore, this chapter first focuses on verifying the actual memory performance. Study 1 was conducted for this question. On the other hand, how to score the different selection of learning strategies could influence this monitoring-control relationship. Accordingly, this chapter also focuses on the verification of scoring the selection of cognitive offloading. Study 2 was conducted to compare the differences between the two scoring methods. In addition, both Studies 1 and 2 are exploratory studies that aim to observe whether specific parameters for the learning tasks could reach clear expectations.

3.1 Introduction for Studies 1 and 2

There are few references for the paired-associate paradigm employed for the current methodology, which focuses on the monitoring-control process at the meta-meta layer via online implementation. Therefore, exploratory research is needed as pilot studies to verify whether some parameter settings are reasonable to reach specific research expectations. In general, the expectations for research settings include three aspects:

1. The difficulty level is sufficient for participants to have their actual memory performance around 0.5, not too high nor too low, prompting the need to consider cognitive offloading in strategy selection.
2. The FOJs are inaccurate in perceiving memory performance, so the SOJs should be needed. This purpose also requires actual memory performance to be neither too high, to

avoid ceiling effects that lead to overestimation, nor too low, to prevent floor effects when compared with FOJs.

3. The cognitive offloading should not be overused to ensure that effort toward memory processes remains adequate.

The detailed aspects of these parameters include:

- The difficulty of paired associates.
- The workload for the learning tasks.
- The learning duration for each learning trial.

Polit studies with initial parameter settings were implemented to verify whether these expectations could be reached.

Among these concerns, the question of the scoring memory performance for strategy choice is critical to the research design. Therefore, Study 2 was implemented to compare the two scoring methods for strategy choice: equally scoring both strategies or scoring the cognitive offloading based on the actual memory performance.

3.2 Study 1

Study 1 examines participants' FOJs, SOJs, and actual memory performance. Specifically, after learning each paired associate, participants promptly responded to a discrete yes/no FOJ scale by indicating whether they believed they could recall the paired associate correctly for the upcoming test (answering "Do you feel you can remember the paired associate correctly for the later test?" by "YES" or "NO"). Then, they responded to the 0-100 numerical (in increments of 10) SOJ scale by indicating their confidence level in the previous FOJ answer in the same trial (answering "Please report how confident you are in the last "YES or NO" answer in the below scale:" by one of the following scales). The instances of employed scales on Gorilla are shown in Figure 9.

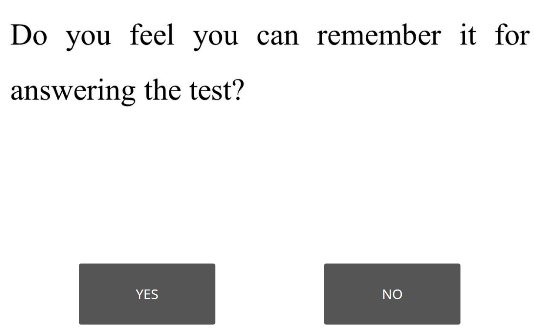


Figure 9A. An instance of the yes/no FOJ scale on Gorilla.

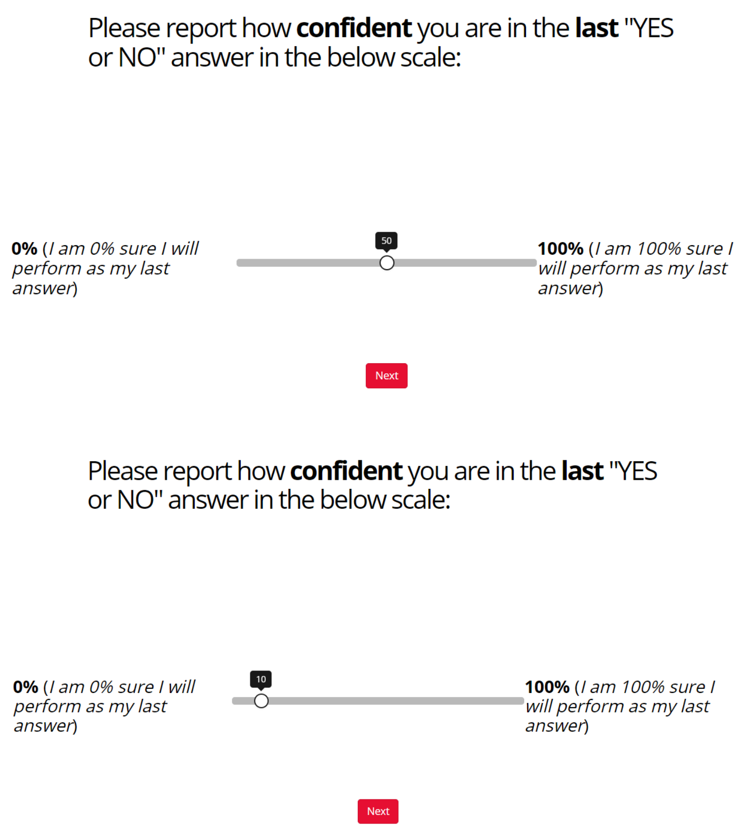


Figure 9B. An instance of the 0-100 numerical scale on Gorilla. The above is the initial state of the scale; the below is an example answer. The participants then click on the “Next” button to submit the answer.

3.2.1 Participants

Thirty-three participants were hired from the Amazon Mechanical Turk (MTurk) platform. Our recruitment was restricted to participants exclusively residing in the United States. Additionally, participation was confined to MTurk workers with a notable reputation, as indicated by a HIT

Approval Rate exceeding 99% (Peer et al., 2014; Peer et al., 2022). The participant sample ranged from 22-35 years old (*M* = 29.06, *SD* = 3.39), with 20 male participants. The average participating duration was 31.24 minutes, with a minimum of 20 minutes and a maximum of 51 minutes. Each participant was compensated with a USD 6 remuneration. Ethical approval had been obtained from the Life Science Committee at the Japan Advanced Institute of Science and Technology (人04-007, from June 15, 2022, to March 31, 2023).

3.2.2 Stimuli and distraction

Fifty-three cue-target English word pairs in uppercase (e.g., ABILITY – CAPABILITY) from South Florida Free Association norms. These pairs exhibit a weak semantic association strength between the cue and target in both forward (0.01) and backward (0) directions, sourced from a database created by Nelson et al. (1998, 2004) for use as learning materials. Of these, five pairs were designated for the Practice session, leaving the remaining 48 pairs reserved for the formal tasks. The participants had five seconds to memorise each pair. The complete list of paired associates is in Appendix 1. For an example of a paired word on Gorilla, see Figure 10.

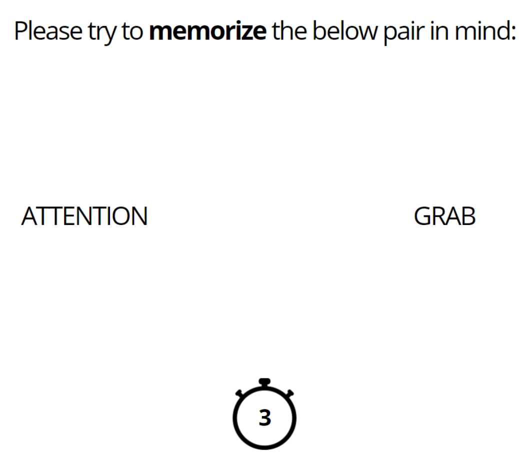


Figure 10. An instance of a paired associate on Gorilla. The left is the cue, and the right is the target. The participants can note the time limitation for learning the presenting pair, indicated by the number on the clock.

An additional session was implemented to mitigate the learning effect between acquiring and recalling paired associates. We used simple arithmetic operations involving pairs of randomly selected numbers from 1-9 for addition and multiplication (144 items). This study does not analyse the scores for this session. The complete list of arithmetic operations is in Appendix 2. For an example of an arithmetic operation on Gorilla, see Figure 11.

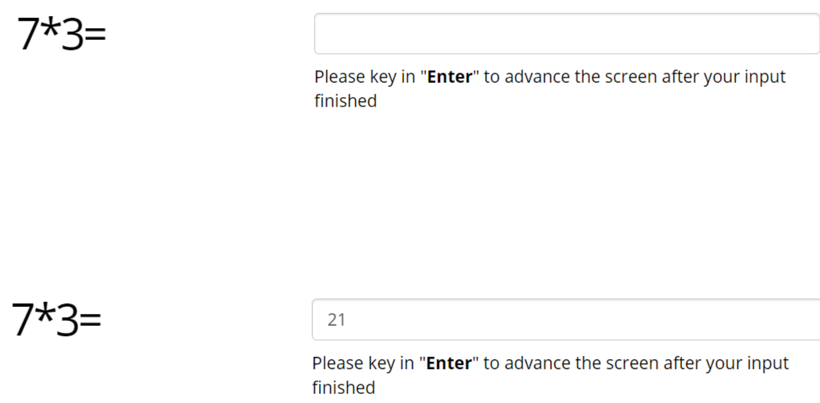
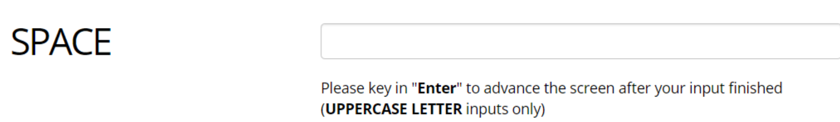


Figure 11. An instance of an arithmetic operation on Gorilla. The above is the initial state. The below is with an input; participants can then key in "Enter" to submit the answer.

3.2.3 Procedure

The experiment comprises three sessions: Learning, Retention, and Test. In the Learning session, the participants learned 48 paired associates randomly. Each pair was displayed on the screen for five seconds per trial, followed by an FOJ and an SOJ. In the Retention session, the participants completed the arithmetic operations. In the Test session, the participants recalled their learned pairs (Figure 12). They did so by referencing the cues provided randomly and inputting the corresponding targets on their keyboards without any time limitations. Additionally, we shortened this procedure to include five pairs and arithmetic operations, creating the Practice session. A general procedure for one trial in the Learning session is shown in Figure 13.



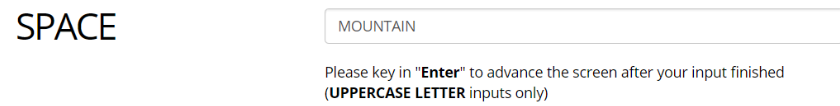


Figure 12. An instance of a recalled paired associate in the Test session on Gorilla. The above is the initial state. The below is with an input; participants can then key in "Enter" to submit the answer.

Figure 13. The procedure for one trial in the Learning session.

During the implementation of this program on MTurk, participants provided their informed consent (Appendix 3), read instructions (Appendix 4), submitted personal information, confirmed adherence to the online participating honour code, and undertook the Practice session.

3.2.4 Results

To determine the difficulty level of the employed paired associates, participants' memory performance in Study 1 was analysed. Accuracy was measured by calculating the number of correct answers during the Test session divided by the total number of trials. This analysis included 33 participants and 48 English word pairs, resulting in 1,584 trials and 694 correct answers, with an overall accuracy of 0.44. The participants correctly answered 21.03 trials (*SD* = 13.22), and each paired associate had an average of 14.46 correct answers (*SD* = 5.55). These findings indicate that these paired associates are moderately challenging and well-suited for the research design.

Study 1 investigated whether participants accurately perceived their memory performance by analysing the consistency between their FOJs and actual memory performance on paired-associate tasks. In Study 1, 1,584 FOJs were collected from participants, with 1,119 indicating they believed they had correctly remembered the paired associate. The average proportion of YES-FOJs was calculated to assess the consistency between perceived and actual memory performance. The findings suggest that participants were overconfident in their memory performance, as the proportion of YES-FOJs (0.71, SD = 0.27) was greater than the overall accuracy of memory performance (0.44). Specifically, the participants reported an average of 33.91 YES-FOJs (SD = 13.01), while their average correct memory performance was 21.03 trials (SD = 13.22), indicating a tendency toward overconfidence in their memory performance. For a detailed confusion matrix, see Table 4.

Table 4. The confusion matrix for illustrating the trial-by-trial comparison between perceived and actual memory performance in Study 1.

	All trials		YES-trials		NO-trials	
	Positive	Negative	Positive	Negative	Positive	Negative
Positive	547	147	547	0	0	147
Negative	572	318	572	0	0	318

The "Positive" and "Negative" for each column represent the predicted condition, and the "Positive" and "Negative" for each row represent the actual condition. The confusion matrix includes the data for all trials, YES-trials, and NO-trials.

Study 1 examined whether participants' confidence level was associated with their perceived memory performance. The findings suggested that all trials' median and mode SOJs were 60, whereas the average was 63.81 (SD = 21.89). The average SOJ was accompanied by an SD of 13.36 at the individual level. The correlation coefficient between FOJs and SOJs calculated via Goodman and Kruskal's gamma was 0.07 ($p < 0.01$), suggesting that FOJs and SOJs maintain independence in memory performance.

The Gorilla platform recorded the duration in milliseconds for each operation to respond to an FOJ (M = 2059.72, SD = 10003.5) and an SOJ (M = 2486.43, SD = 3175.66). The time durations for FOJs in the YES trials (M = 1693.32, SD = 2140.13) were significantly shorter than those in the NO-trials (M = 2941.42, SD = 18145.63), as indicated by an independent samples t-test (t (1528) = 2.26, p = 0.02). Spearman's correlation revealed a significant positive correlation between the durations of FOJ responses and SOJ responses ($\rho = 0.21$, $p < 0.01$), indicating that FOJs consuming longer durations are frequently preceded by SOJs exhibiting similarly extended durations.

On the other hand, a chi-squared test revealed a significant difference between perceived and actual memory performance ($\chi^2(1) = 231.64$, $p < 0.01$). Related contingency table see Table 5.

Table 5. The contingency table between perceived and actual memory performance.

	Can remember	Cannot remember	Total trials
Perceived memory	1119	466	1585
Actual memory	694	891	1585

The "Can remember" column represents the number of trials either with YES-FOJ for perceived memory performance or with correct answers for actual memory performance. The "Cannot remember" column represents the number of trials either with NO-FOJ for perceived memory performance or with incorrect answers for actual memory performance.

3.2.5 Conclusions of Study 1

Study 1 aimed to test the main study's feasibility of introducing cognitive offloading as a strategy. The difficulty level of the selected paired associates was evaluated to ensure that the participants faced a moderate challenge. The analysis of participants' accuracy in recalling the paired associates yielded a score of 0.44, indicating that the selected paired associates have the potential for improvement with cognitive offloading aid. These findings are within the expectation for the difficulty level of learning tasks, suggesting that participants may benefit from introducing cognitive offloading to enhance their performance.

Study 1 also examined whether participants could accurately perceive their memory performance. The consistency between their perceived and actual memory performance indicated the FOJ accuracy. The results showed that most participants were overconfident in their memory performance, which is consistent with previous research (Rhodes, 2016). These findings reached expectations regarding the accuracy of FOJs, suggesting that participants may fail to select an appropriate strategy based on their FOJs during practice. Furthermore, these findings highlighted the implications of investigating the relationship between metacognition and cognitive offloading (Risko & Dunn, 2015; Dunn & Risko, 2016), suggesting that assessments from high-order metacognitive judgements (i.e., meta-meta layer) rather than perceived memory performance may be needed. A comparison of significant differences ($\chi^2(1) = 231.64$, $p < 0.01$) between the actual and perceived memory performance is shown in Figure 14.

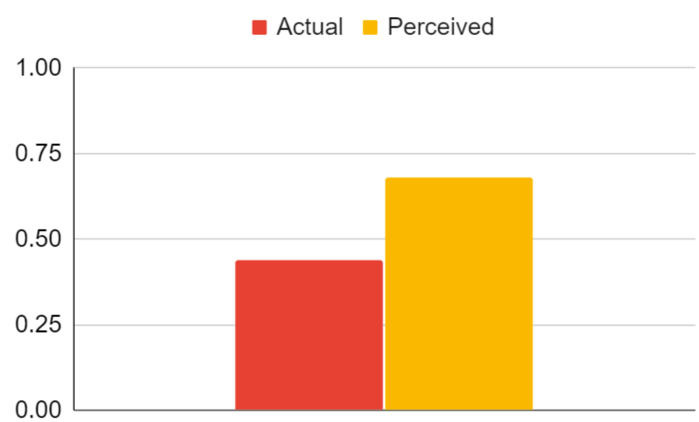


Figure 14. The comparison between the actual (red) and perceived (yellow) memory performance.

3.3 Study 2

The independent variable of this experiment includes two scoring methods for different strategy choices – using cognitive offloading or not. Specifically, two conditions in this research were as follows: (1) consider the correct answers in the Test session as equal performance (Study 2A), or (2) consider the performance of the correct answers based on the cognitive offloading strategy to be only half of these based on memory (Study 2B), given that this weight (0.5) is close to the mathematical expectation, considering the correctness should around 0.4 to 0.5 (with 0.44 being suggested in Study 1). For this purpose, Study 2 incorporates the selection of strategy choice into the paired-associate paradigm. Related details are given in the following subsections.

3.3.1 Participants

Sixty participants were hired from the MTurk platform. This recruitment was restricted to participants exclusively residing in the United States. Additionally, participation was confined to MTurk workers with a notable reputation, as indicated by a HIT Approval Rate exceeding 99%

(Peer et al., 2014; Peer et al., 2022). The participant sample ranged from 22-44 years old (*M* = 30.00, *SD* = 3.67), with 37 male participants. The average participating duration was 39.13 minutes, with a minimum of 27 minutes and a maximum of 58 minutes. Each participant was compensated with a USD 6 remuneration. Ethical approval had been obtained from the Life Science Committee at the Japan Advanced Institute of Science and Technology (人04-007, from June 15, 2022, to March 31, 2023).

Two studies were implemented with a between-subject design. The participants were randomly assigned to Study 2A (*n* = 32) or Study 2B (*n* = 28) with a probability of 0.5.

3.3.2 Tasks, Materials, and Environment

This study used identical paired associates sourced from South Florida Free Association norms as those employed in Study 1. As the allowance of cognitive offloading, one additional question, “Please select your strategy between ‘Writing down’ and ‘Keeping in Mind’:”, was included in the experiment, asking participants to choose between two strategies, with two options: “Writing down” and “Keeping in mind,” to determine whether they offload or not (for an instance of cognitive offloading on Gorilla, see Figure 15). The same online platforms used in Study 1 were used for this implementation.

Figure 15A. An instance of selecting learning strategies on Gorilla. If participants select “Keeping in Mind,” they finish this trial and advance to the next paired word.

Figure 15B. The instance of selecting learning strategies on Gorilla when “Writing down” is selected. The participants can write the paired word for later reference without time limitations and click “Written down” to continue the program.

3.3.3 Procedure

The procedure for Study 2 closely resembled that described in Study 1. One exception was that participants were asked to report their strategies during the Learning session immediately after submitting each SOJ. If they employed the “Writing down” strategy, they were shown the pair again without a constrained time frame and could write it down on paper for reference during the Test session. Participants in the Test session of Study 2A received 1 point for each correct answer, regardless of the strategy they selected for that trial. In the Test session of Study 2B, participants received 2 points for a correct answer from memory, 1 point for a correct answer from the paper, and 0 points for an incorrect answer. The goal of this experiment was not to analyse the score of each participant. For the Learning session procedure, see Figure 16. Additionally, for the procedures of all the sessions, see Figure 17.

Figure 16. The procedure for one trial in the Learning session. The “Offloading” screen will appear depending on the “Writing down” strategy selected in the “Strategy selection.”

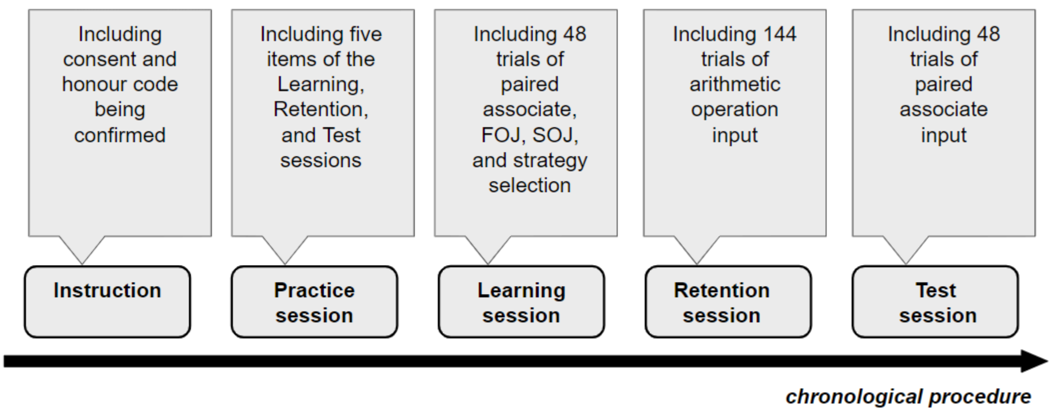


Figure 17. The chronological procedure for the sessions is outlined below, along with the contents included in each session above.

3.3.4 Results

Study 2A included a total of 1,536 trials involving 32 participants. Among them, 1,305 YES-FOJs were received, 810 responses were received with the offloading strategy, and the average SOJ was 80.23 (*SD* = 20.00). On average, each participant reported that they could remember 40.78 (*SD* = 12.76) pairs, and they selected 25.31 (*SD* = 20.60) offloading. On average, each pair responded YES-FOJ 27.19 (*SD* = 1.82) times and was offloaded 16.88 (*SD* = 2.14) times.

Study 2B included a total of 1,344 trials involving 28 participants. Among them, 1,112 YES-FOJs were received, 513 responses with the offloading strategy, and an average SOJ of 72.50 (*SD* = 20.91). On average, each participant reported that they could remember 39.71 (*SD* = 10.96) pairs, and they selected 18.32 (*SD* = 17.64) offloading. On average, each pair responded YES-FOJ 23.17 (*SD* = 2.37) times and was offloaded 10.69 (*SD* = 2.57) times. The ratio between trials employing cognitive offloading and total trials is 0.53 for Study 2A and 0.38 for Study 2B. On another note, while there were 725 trials employing cognitive offloading among 1,305 trials of YES-FOJ in Study 2A (a ratio of 0.56), there were 386 trials employing cognitive offloading among 1,112 trials of YES-FOJ in Study 2B (a ratio of 0.35 that significantly lower than the ratio of 0.56 in Study 2A, χ^2 (1) = 104.18, $p < 0.01$, for contingency table see Table 6). A comparison is shown in Figure 18. These results suggested that considering the correct answers in the Test session and employing cognitive offloading, as half of these are based on memory, can reduce the selection of cognitive offloading compared with considering the correct answers in the test session, with both strategies being equal.

Table 6. The contingency table between strategy choices of Studies 2A and 2B.

	Writing down	Keeping in mind	Total trials
Study 2A	725	580	1305
Study 2B	386	726	1112

The columns represent the number of trials with related learning strategies. The rows represent the number of trials in related conditions of Study 2.

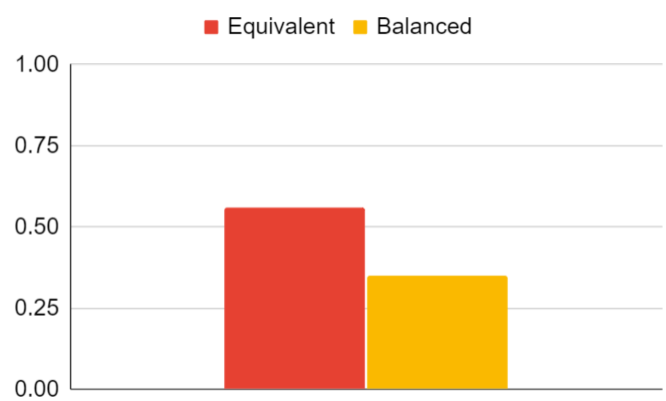


Figure 18. The comparison of trials employing cognitive offloading between equivalent (red) and balanced (yellow) scoring ways.

3.4 Summary

In this chapter, two pilot studies were conducted to verify various parameters in the research setting for the main study with the paired-associated paradigm. First, a relatively low correct rate of actual memory performance and overconfidence in FOJs were observed, which aligns with expectations that the materials were appropriate regarding the difficulty of remembering the paired associates, the workload for the learning tasks, and the learning duration for each learning trial. Second, Study 2 suggests that considering the correct responses in the Test session by employing cognitive offloading, as half of these are based on memory, can reduce the reliance on the cognitive offloading strategy. The subsequent main study was designed in line with these findings, as demonstrated in Chapter 4.

Chapter 4 Main Study

Overview

Study 3 (the main study) continued the data collection of Study 2B, as the research settings in Study 2B fit the aim of this dissertation – exploring the interaction between FOJs and SOJs on the selection of cognitive offloading. This chapter reported details regarding this analysis with a larger sample size. The aimed relationship was subsequently verified by fitting linear mixed-effects models, considering FOJs, SOJs, and their interaction term as fixed effects and individual participants and paired-associate items as random effects. Most of the content introduced in this chapter has been published in the journal *Discover Psychology* under the title "The influence of second-order metacognitive judgements on cognitive offloading within the monitoring-control relationship" through a peer-review process under a licence of CC BY 4.0 (Ma & Fujinami, 2024).

4.1 Participants

Study 3 was built on Study 2B by recruiting 52 more participants using the same criteria. Eventually, 80 participants (49 male; age: $M = 33.44$, $SD = 4.0$, ranging from 24-44) were recruited from the United States through Amazon Mechanical Turk (MTurk). The participants were restricted to those with a high reputation record (HIT approval rate > 99%) (Peer et al., 2014; Peer et al., 2021). The average participation time was 41.19 minutes with a minimum of 25 minutes and a maximum of 87 minutes, and the participants received a reward of USD 6. All participants provided their consent and agreed to an honour code. They were informed that their participation would be rejected, their MTurk reputation would be reduced, and their reward would be cancelled if their behaviour exceeded the instructions.

4.2 Descriptive statistics

The participants' responses were recorded, including their first-order metacognitive judgements (FOJs) on the yes/no scale, their second-order metacognitive judgements (SOJs) on the 0-100 scale, and their strategy choice reports. A total of 3,840 trials were received from 80 participants (48 trials per person). Precisely, 3,142 trials with a "YES" answer for FOJs, 1,563 responses employing cognitive offloading, and an average SOJ of 73.59 (*SD* = 21.60) were received. Table 7 presents a contingency table of FOJs and strategy choices. On average, each participant reported that they could remember 39.26 (*SD* = 12.80) pairs and offloaded 19.53 (*SD* = 17.25) times. On average, each paired associate received 66.83 (*SD* = 10.13) "YES" responses to the FOJ and was offloaded 33.23 (*SD* = 7.97) times. The participants' performance in the Retention session averaged 137.21 (*SD* = 13.92) correct trials out of 144 trials. The participants' performance in the Test session averaged 33.94 (*SD* = 13.76) correct trials out of 48 and 52 (*SD* = 22.22) points out of a maximum possible score of 96. The participants spent an average of 15.58 minutes (*SD* = 4.36) in the Learning session, 7.79 minutes (*SD* = 2.52) in the Retention session, and 6.93 minutes (*SD* = 3.27) in the Test session.

Table 7. Contingency table of FOJs and strategy choices.

	"Writing down" Strategy	"Keeping in Mind" Strategy	Total
YES-FOJ	1203	1939	3142
NO-FOJ	359	339	698
Total	1562	2278	3840

4.3 Distributions

The frequencies for each SOJ level and the average per person for each SOJ level were distributed, as shown in Figures 19 and 20. Anderson-Darling (AD) tests were implemented to verify the goodness of fit, which showed a good fit with a three-parameter Weibull distribution (AD = 0.42, $p = 0.35$), or a good fit with a normal distribution after being processed by the Johnson transformation (AD = 0.24, $p = 0.75$). For more results, see Table 8. Moreover, the probability plots for frequency for more distributions are shown in Figure 21, in which the points on a plot following a straight line within the confidence bounds indicate the good fit for data or the effective transformation. The frequency of cognitive offloading per trial averaged 32.54 (*SD* = 2.62).

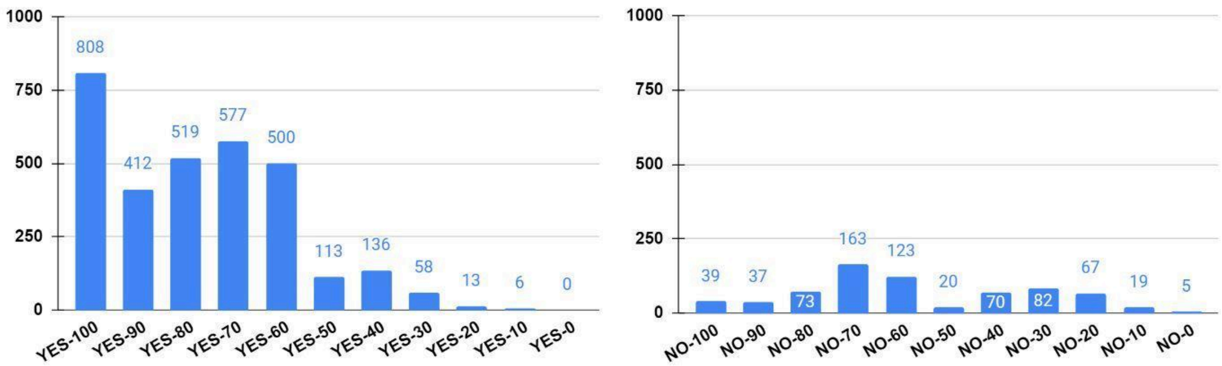


Figure 19. The frequencies correspond to each SOJ level for FOJs with a YES answer (left) and a NO answer (right). The horizontal axis represents each SOJ category, whereas the vertical axis represents the frequency of reported occurrences within each category.

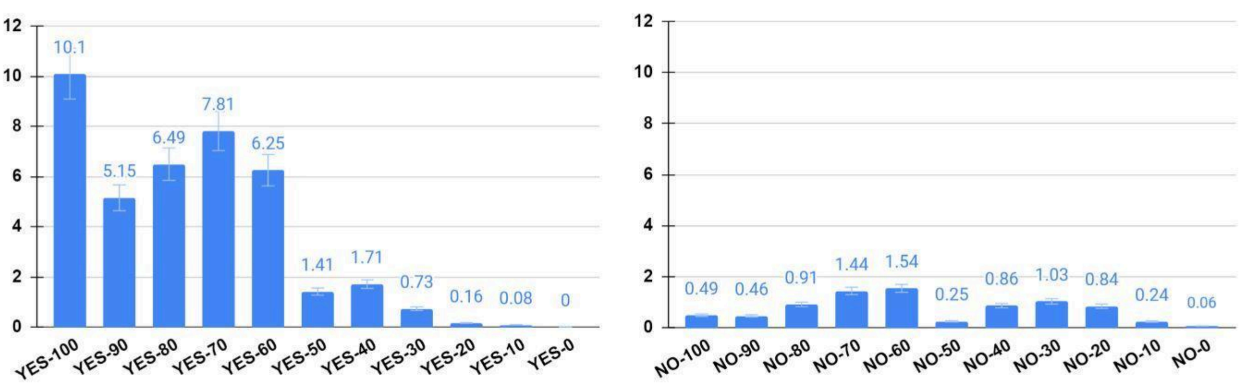
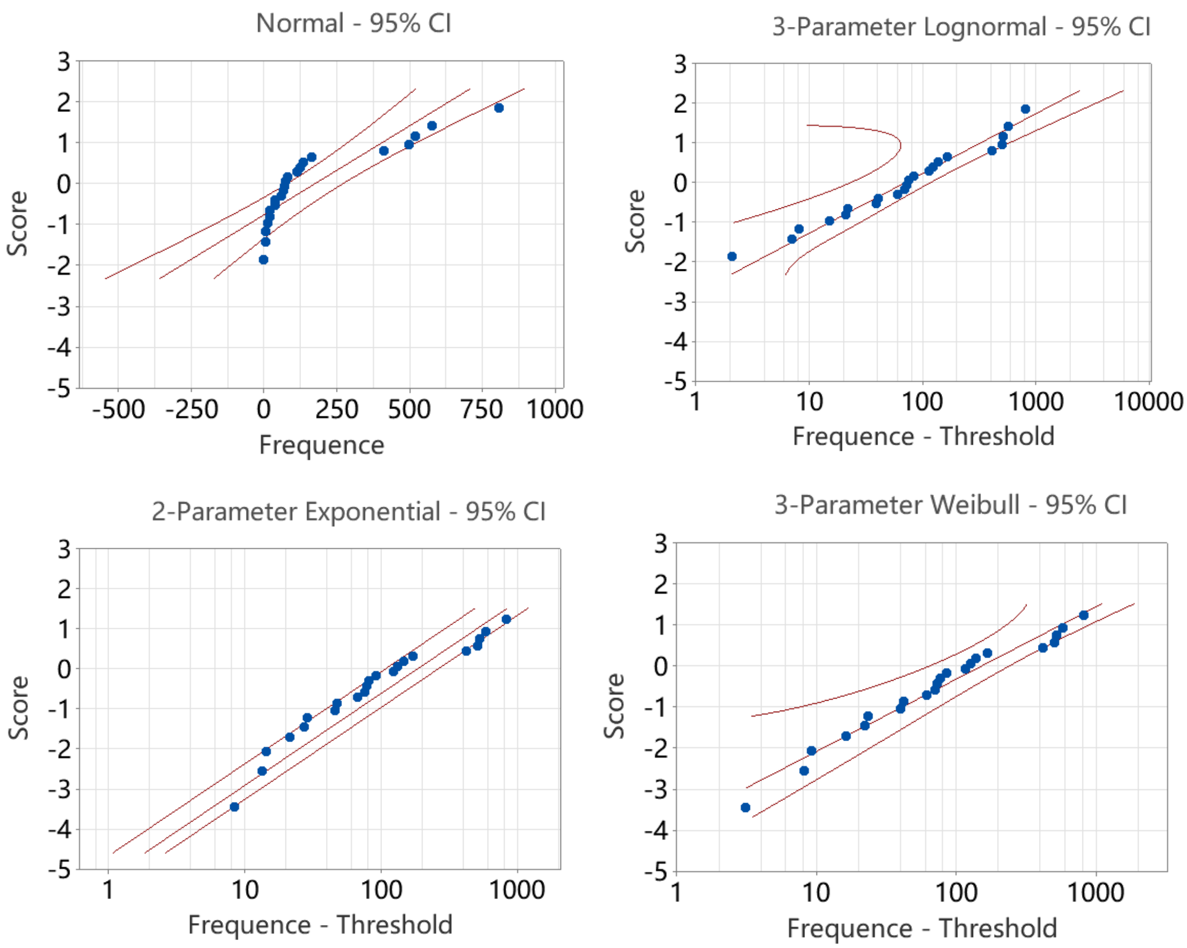


Figure 20. The frequencies per person correspond to each SOJ level for FOJs with a YES answer (left) and a NO answer (right), averaged across participants. The horizontal axis represents each SOJ category, whereas the vertical axis represents the average frequency of reported occurrences within that category.

Table 8. The goodness of fit test for the distribution of cognitive offloading frequency.

Distribution	Anderson-Darling tests	*p*-value
Normal	2.54	< 0.01
3-parameter log-normal	0.26	-
2-parameter exponential	1.32	0.02
3-parameter Weibull	0.42	0.35
Smallest extreme value	2.80	< 0.01

Largest extreme value	2.07	< 0.01
3-parameter gamma	0.50	-
Logistic	2.30	< 0.01
3-parameter log-logistic	0.28	-
Johnson transformation	0.24	0.75



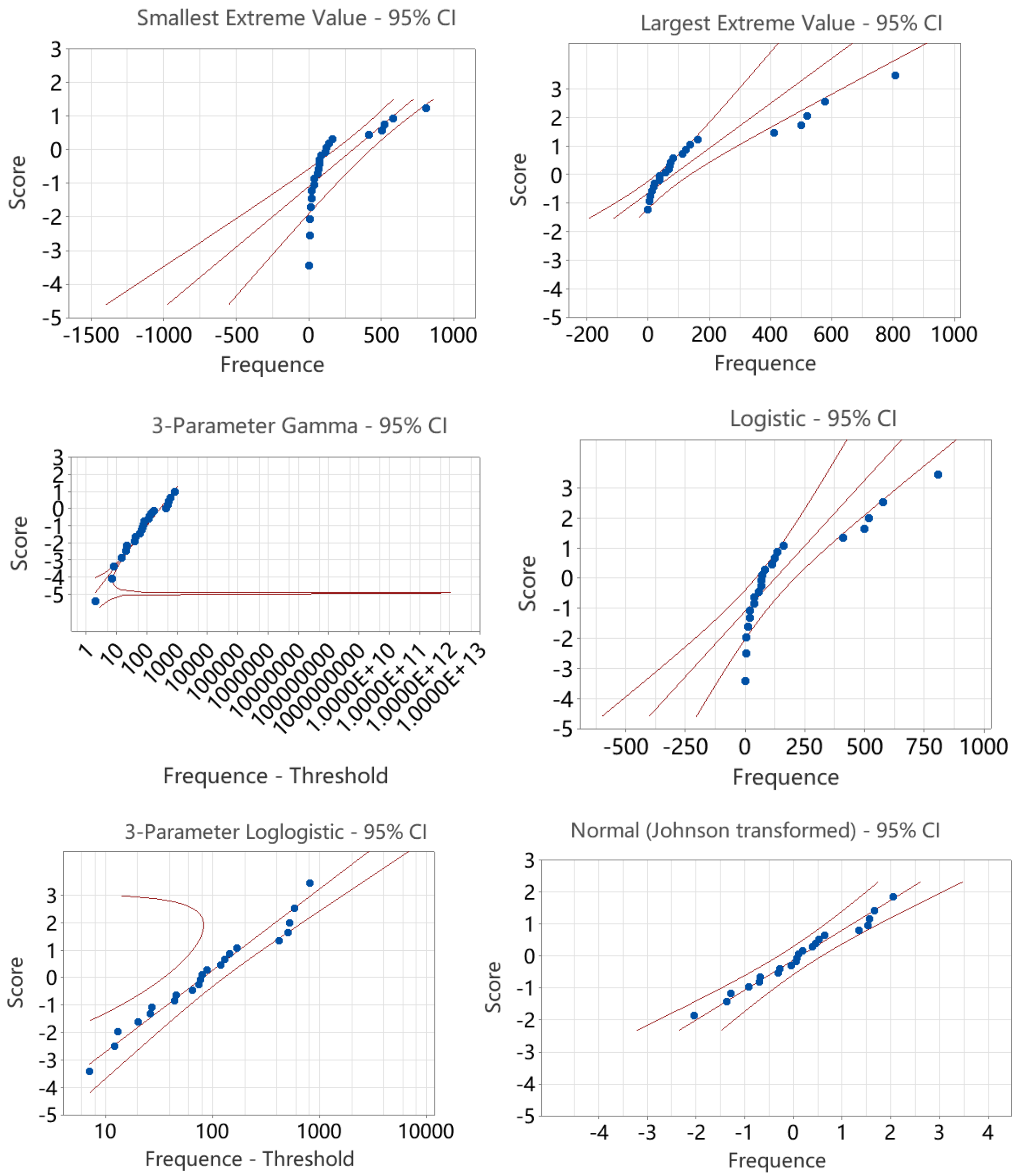


Figure 21. The probability plots from the data for various distributions with a confidential interval of 95% (95% CI). The horizontal axis represents the frequency values, whereas the vertical axis represents the inverse cumulative probabilities.

The distribution by trial is shown in Figure 22. A one-way ANOVA test revealed no significant difference (F (2, 45) = 0.50, p = 0.61) among the first third of the trials (Trial 1-16, M = 32.06,

SD = 3.34), the second third of the trials (Trial 17-32, M = 33.00, SD = 1.37), and the final third of the trials (Trial 33-48, M = 32.56, SD = 2.83), suggesting that cognitive offloading was employed equally across the learning span. Levene's test for equality of variances confirmed that the homogeneity of variance assumption was marginally significant (F (2, 45) = 3.27, p = 0.05). A linear regression was conducted with trial order as a trial order-based covariate and cognitive offloading frequency as the dependent variable, revealing no significant relationship between them (F (46) = 0.45, p = 0.51). These results suggest that the cognitive offloading frequency was consistent across the 48 trials.

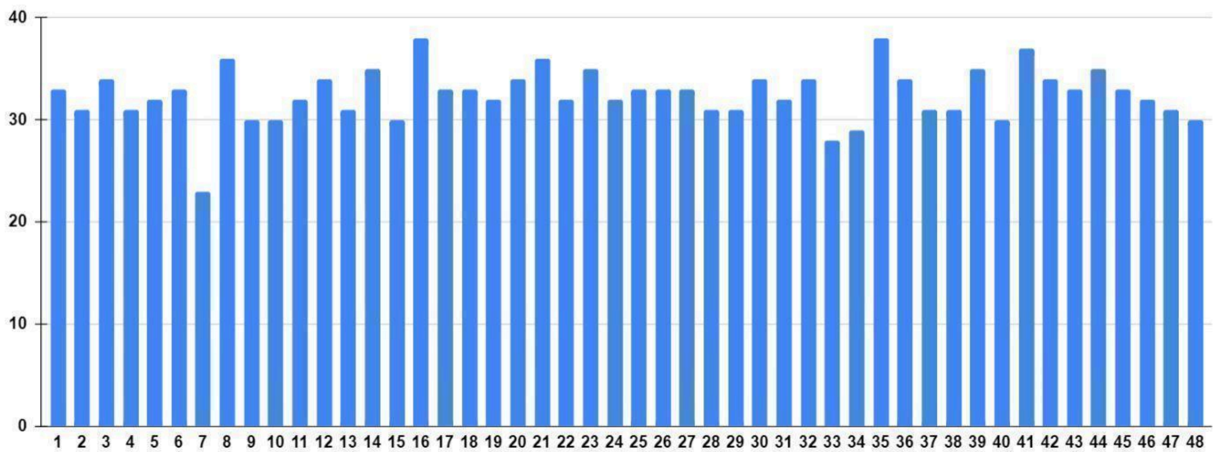


Figure 22. The average frequency of cognitive offloading employed across 48 trials. The horizontal axis represents each trial from 1-48, and the vertical axis represents the frequency of cognitive offloading.

4.4 Correlations

Goodman and Kruskal's gamma was 0.53 (p < 0.01) for the correlation between FOJs and SOJs. The related contingency table is presented in Table 9. This finding suggests a moderate positive correlation between the two variables, alleviating concerns about multicollinearity affecting the FOJs and SOJs.

Table 9. Contingency table for the FOJ-SOJ relationship.

	SOJs										
FOJs	0	10	20	30	40	50	60	70	80	90	100	Total

YES	0	6	13	58	136	113	500	577	519	412	808	3142
NO	5	19	67	82	70	20	123	163	73	37	39	698
Total	5	25	80	140	206	133	623	740	592	449	847	3840

Spearman's correlations were calculated between individuals' correct trials in the Test session and their FOJs, SOJs, and strategy choices. A significant correlation ($\rho = 0.32$, $p < 0.01$) was found between participants' correct trials in the Test session and the frequency of YES-FOJ, suggesting that participants recalled more trials when they felt they could remember the paired associates. A significant correlation ($\rho = 0.42$, $p < 0.01$) was found between participants' correct trials in the Test session and their average SOJs, suggesting that participants recalled more trials when their confidence levels were higher. A significant correlation ($\rho = 0.4$, $p < 0.01$) was found between participants' correct trials in the Test session and the frequency of the "Writing down" strategy, suggesting that those who recalled more trials in the Test session employed cognitive offloading strategies more frequently.

Spearman's correlations between participants' strategy choices and their time spent in the Learning, Retention, and Test sessions were calculated. A significant correlation ($\rho = 0.42$, $p < 0.01$) was found between participants' frequency of the "Writing down" strategy and their time spent in the Learning session, suggesting that those who employed cognitive offloading more frequently completed the Learning session more slowly. An insignificant correlation ($\rho = 0.17$, $p = 0.13$) was found between participants' frequency of the "Writing down" strategy and their time spent in the Retention session, suggesting that the choice between "Keeping in mind" and "Writing down" was unrelated to the efficiency of completing the arithmetic operations. A significant correlation ($\rho = 0.35$, $p < 0.01$) was found between participants' frequency of the "Writing down" strategy and their time spent in the Test session, suggesting that those who employed more cognitive offloading took longer to input the paired associates.

A post hoc power analysis for Spearman's correlation was conducted via G*Power to determine the statistical power (Faul et al., 2009), given a sample size of 80 and an alpha error probability of 0.05. The analysis yielded a power of 0.22. This result may support the argument that insufficient sample size is due to the insignificant correlation between participants' frequency of the “Writing down” strategy and their time spent in the Retention session.

4.5 Generalised linear mixed models

The relationships between FOJs, SOJs, and strategy choices across individual participants and paired associates were investigated via generalised linear mixed models. First, the degree of variability explained solely by random effects was assessed. A random effect for individual participants accounted for 52.87% of the total variance ($VAR = 0.13$, $p < 0.01$). Additionally, a random effect for paired associates accounted for 1.19% of the total variance ($VAR = 2.88\text{e-}3$, $p < 0.01$). These findings suggest that individual participants and paired associates significantly contribute to the variability in strategy choices, indicating their importance in the relationships between FOJs, SOJs, and strategy choices. The formulas for these models are shown in Table 10 (Nos. 1 and 2).

Table 10. Specifications of three different generalised linear mixed models used to analyse data on strategy choices.

No	Formula
1	strategy choices ~ 1 + (1 | individual participants)
2	strategy choices ~ 1 + (1 | paired associates)
3	strategy choices ~ FOJs + SOJs_s + FOJs * SOJs_s + (1 | individual participants) + (1 | paired associates)
4	strategy choices ~ FOJs + (1 | individual participants) + (1 | paired associates)
5	strategy choices ~ FOJs + SOJs_s + (1 | individual participants) + (1 | paired associates)

Formula 1 includes a random intercept for individual participants. Formula 2 includes a random intercept for paired associates. Formula 3 includes fixed effects for FOJs, standardised SOJs (denoted as SOJs_s), and their interaction (denoted as FOJs*SOJs_s), with random intercepts for individual participants and paired associates. Formula 4 is a nested model of Formula 3, considering the coefficients of SOJs_s and FOJs*SOJs_s are 0. Formula 5 is a nested model of Formula 3, considering the coefficient of FOJs*SOJs_s is 0.

Second, the FOJs, SOJs, and strategy choices were preprocessed for model fitting. The yes/no scale for FOJs was converted to numerical values, with "YES" coded as "1" and "NO" coded as "0." The 0-100 scale for SOJs was standardised by subtracting the mean from each value and dividing by the standard deviation. The interaction term was then calculated by multiplying the converted FOJs with the standardised SOJs. For strategy choices, "Writing down" was coded as "1," and "Keeping in mind" was coded as "0."

Finally, FOJs, SOJs, and their interactions were included as fixed effects, with individual participants and paired associates included as random effects, and strategy choices were included as the dependent variable. The random effects for individual participants and paired associates accounted for 57.97% ($VAR = 0.14$, $p < 0.01$) and 0.81% ($VAR = 2.03\text{e-}3$, $p < 0.01$) of the variance, respectively. FOJs had a significant negative main effect ($\beta = -0.35$, $p < 0.01$), indicating that the likelihood of employing cognitive offloading is lower when the FOJ response is "YES" than when it is "NO." Additionally, SOJs had a significant positive main effect ($\beta = 0.04$, $p < 0.01$), indicating that the likelihood of employing cognitive offloading is greater with a greater SOJ. Finally, a significant interaction effect between FOJs and SOJs was found ($\beta = -0.10$, $p < 0.01$), indicating that the relationship between FOJs and strategy choices depends on SOJs. The negative coefficient ($\beta = -0.10$) suggests that the negative effect of FOJs on the likelihood of employing cognitive offloading becomes more pronounced when the SOJ is higher. The model accounted for 58.53% of the variation, reducing to 58.50% after adjustment. The formula for this model is in Table 10 (No. 3). The details of this model are presented in Table 11. The interaction plot from this model is shown in Figure 23, where the slopes of the regression lines represent the relationships of FOJs and SOJs with the selection of cognitive offloading, each depicted in a distinct colour, and the nonparallel lines suggest an interaction effect between FOJs and SOJs on the selection of cognitive offloading.

Table 11. Summary of fixed and random effects from the generalised linear mixed model (Formula 3).

	Effect	Estimate/ variance	Standard error	*t* value/*z* value	*p* value
Fixed effects	Constant	0.70	0.05	14.70	< 0.01
	FOJs (converted)	-0.35	0.02	-15.24	< 0.01

	SOJs (standardised)	0.04	0.01	3.03	< 0.01
	Interaction term	-0.10	0.02	-5.97	< 0.01
Random effects	Individual participants	0.14	0.02	6.17	< 0.01
	Paired associates	2.03e-3	7.00e-4	2.91	< 0.01

In the column "Estimate/variance," the "estimate" is for the fixed effects, and the "variance" is for the random effects. In the column "*t* value/*z* value," the "*t* value" is for the fixed effects, and the "*z* value" is for the random effects.

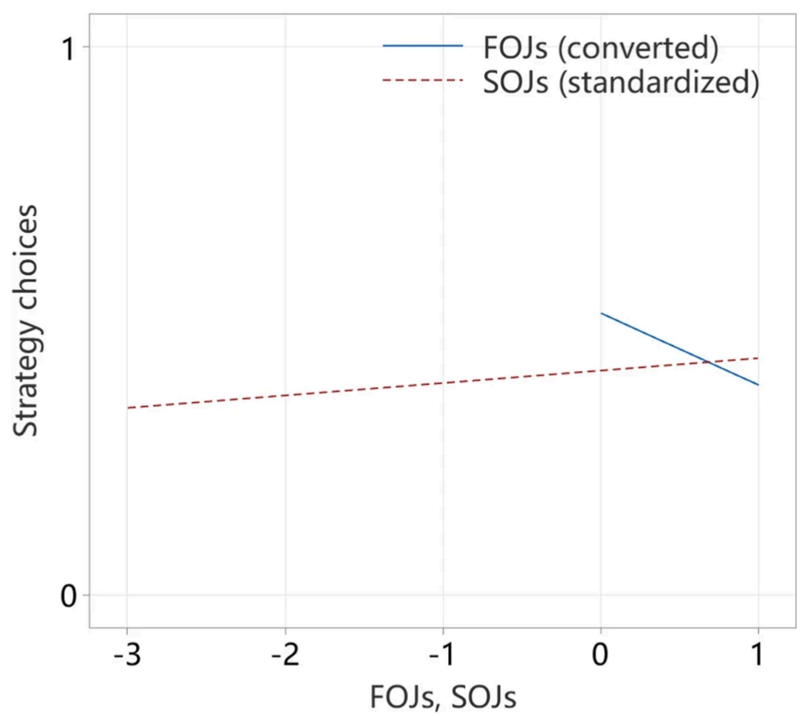


Figure 23. Interaction plot from the generalised linear mixed model (Formula 3).

In addition, nested models of the generalised linear mixed model (Formula 3) were fitted to compare model performance after removing fixed effects. Formula 4 (No. 4 in Table 10) included FOJs as fixed effects, with individual participants and paired associates included as random effects, and strategy choices were included as the dependent variable. The random effects for individual participants and paired associates accounted for 56.19% ($VAR = 0.14$, p

< 0.01) and 1.11% ($VAR = 2.71\text{e-}3$, $p < 0.01$) of the variance, respectively. FOJs had a significant negative correlation with strategy choices ($\beta = -0.28$, $p < 0.01$), indicating that the likelihood of employing cognitive offloading is lower when the FOJ response is "YES" than when it is "NO."

Formula 5 (No. 5 in Table 10) included FOJs and SOJs (standardised) as fixed effects, with individual participants and paired associates included as random effects, and strategy choices were included as the dependent variable. The random effects for individual participants and paired associates accounted for 57.05% ($VAR = 0.14$, $p < 0.01$) and 0.98% ($VAR = 2.42\text{e-}3$, $p < 0.01$) of the variance, respectively. FOJs had a significant negative correlation with strategy choices ($\beta = -0.27$, $p < 0.01$), indicating that the likelihood of employing cognitive offloading is lower when the FOJ response is "YES" than when it is "NO." SOJs had a significant negative correlation with strategy choices ($\beta = -0.03$, $p < 0.01$), indicating that higher SOJs were associated with less frequent selection of cognitive offloading. A comparison of model performance across all formulas is seen in Table 12. Among them, Formula 3 has the highest R-squared and adjusted R-squared values, indicating that it explains the most variance in the dependent variable. Also, Formula 3 has the lowest Akaike's corrected information criterion and Bayesian information criterion values, suggesting it is the most parsimonious model with the best balance of fit and complexity.

Table 12. Comparison of model performance across all formulas.

Formula	R-sq	R-sq (adj)	AICc	BIC
No. 1	53.52%	53.52%	2903.60	2916.11
No. 2	1.76%	1.76%	5436.71	5449.21
No. 3	58.53%	58.50%	2580.02	2598.78
No. 4	58.08%	58.07%	2613.06	2631.81
No. 5	58.19%	58.17%	2608.90	2627.65

R-sq denotes the R-squared value, referring to the percentage of variation in the response that is explained by the model. R-sq (adj) denotes the adjusted R-squared value, referring to a modified version of R-squared that accounts for the number of predictors in the model. AICc denotes Akaike's corrected information criterion, referring to an information criteria used to compare the quality of statistical models for a given set of data. BIC denotes the Bayesian information criterion, penalising model complexity more heavily than AICc.

4.6 Consistency within SOJ levels

A trial with a “YES” FOJ and a “Keeping in mind” strategy (or a “NO” FOJ and a “Writing down” strategy) was defined as a consistent trial. The ratio of consistent trials to the total number of trials within each SOJ was then defined as the consistency for that level. These consistencies are depicted in Figure 24.

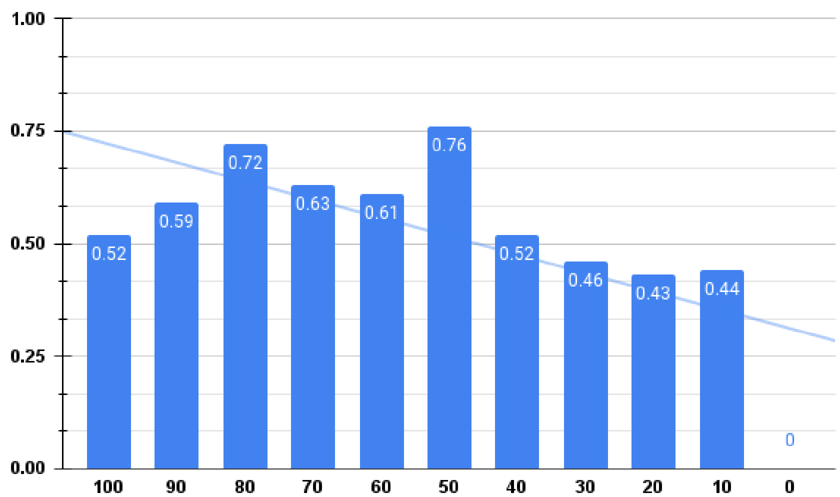


Figure 24. The consistencies for each SOJ level (horizontal axis, ranging from 0-100) in ratios (vertical axis, ranging from 0-1). The blue line indicates the overall trend of the consistency ratios across the SOJ levels.

Chapter 5 Discussion

Overview

This chapter discusses the results obtained from studies in earlier chapters on incorporating second-order metacognitive judgements (SOJs) into the relationship between first-order metacognitive judgements (FOJs) and strategy choices. The content was divided into three subsections. First, the observed FOJ-SOJ interaction was interpreted for monitoring and control processes. Second, the necessity of considering the SOJ as an assessment at the meta-meta layer was argued, alongside an alternative approach for regulating monitoring accuracy and the rationale for the irreplaceability of the SOJ. Finally, limitations in the methodology were acknowledged, focusing primarily on the employed paired-associate paradigm and its remote implementation.

5.1 Interpretation

This dissertation revealed that while FOJs predict strategy choices to some extent, SOJs further moderate the strength of the association between FOJs and the selection of cognitive offloading. Accordingly, a significant FOJ-SOJ interaction in the selection of cognitive offloading can be identified. This interaction reflects two distinct metacognitive processes involved in forming the SOJ, namely, monitoring and control, extending beyond the relationship between the FOJs and the strategy choices.

The monitoring process suggests that SOJ formation involves assessing prior FOJs as monitoring results, gathering evidence regarding the consistency between perceived and actual memory performance. This process relies on self-access to the FOJ, with updates to the SOJ formation depending on whether new evidence is uncovered during each monitoring cycle. The accumulation of such evidence may indicate FOJ accuracy; therefore, when SOJs are explicitly elicited, their magnitude should directly reflect the accuracy of the corresponding FOJ. This monitoring process aligns with the dual perspective proposed by Dunlosky et al. (2005), as illustrated in the model shown in Figure 25.

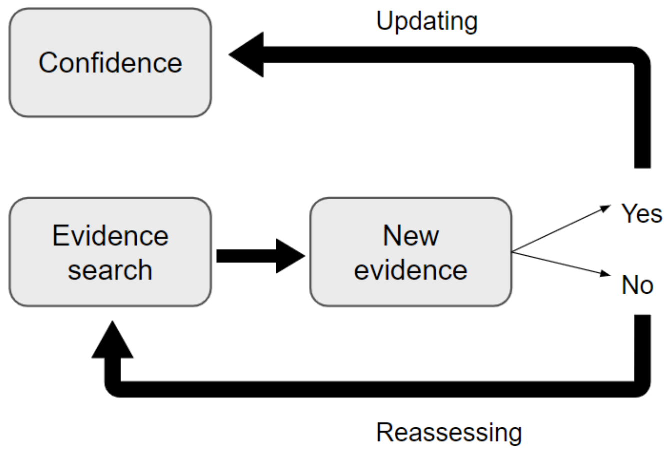


Figure 25. A cognitive model regarding how an SOJ monitors the meta layer monitoring-control relationship. The bold black arrows indicate the direction of the process. The rectangles with rounded corners indicate the cognitive component in the process. The thin arrows indicate the branches in the process.

The control process acknowledges the intuitive association between perceived memory performance and strategy choice (i.e., a pessimistic assessment associated with a conservative strategy) as a precondition. Furthermore, this process suggests that the accumulation of evidence may reverse the intuitive association once the SOJ falls below a specific threshold. For example, if memory performance is perceived as "YES" at the meta layer, the learning strategy should be intuitively associated with "Keeping in mind." However, if its corresponding confidence level at the meta-meta layer is rated as "0" (an extreme case of low confidence), this association could be regulated to the opposite strategy, "Writing down." In this case, the explicitly reported confidence level – namely the SOJ – could enhance the predictive power of learning strategies by compensating for an inaccurate FOJ. This process interprets how the SOJ regulates the FOJ to enhance monitoring accuracy, as illustrated in Figure 26.

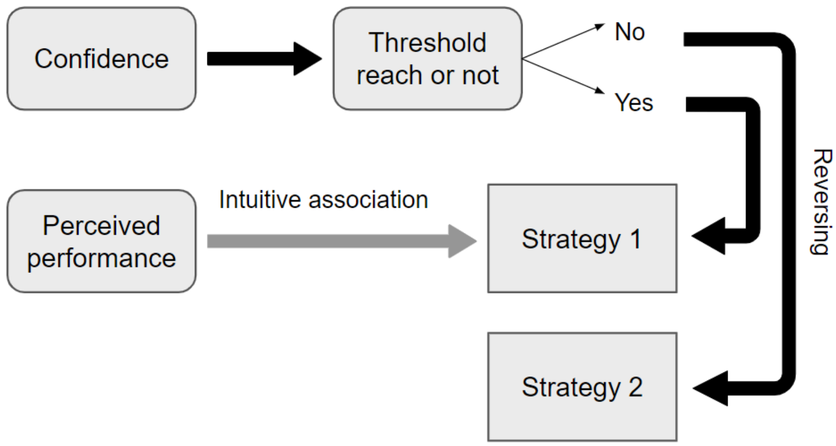


Figure 26. A cognitive model regarding how an SOJ controls the monitoring-control relationship at the meta layer. The bold grey arrows indicate intuitive associations without the control process involved. Rectangles with sharp corners indicate observable behaviours.

In summary, this dissertation incorporates the SOJs as eliciting a higher-order metacognitive process over the monitoring-control relationship as a meta-meta layer beyond the meta layer. This perspective underscores the dual-process hypothesis, emphasising how SOJ enhances monitoring accuracy, as proposed by Dunlosky et al. (2005). On the one hand, it initially illustrates how control adjusts strategy choices based on confidence of SOJ thresholds. The new monitoring and control processes are summarised in Figure 27 and are illustrated with detailed processes as a simplified diagram within Nelson and Naren's (1990) metacognitive architecture in Figure 28.

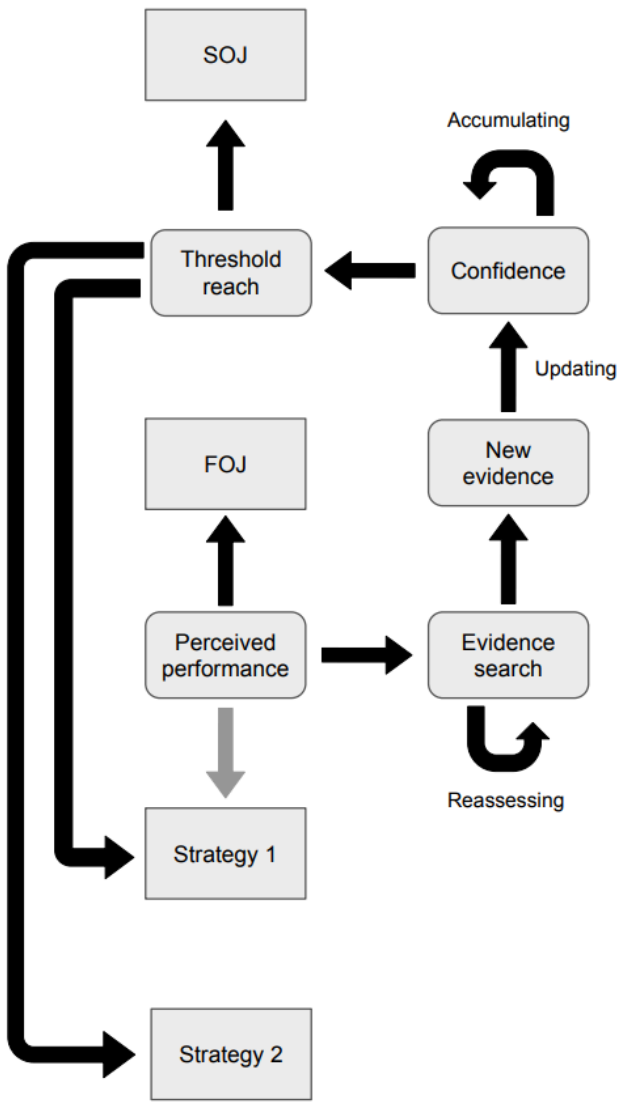


Figure 27. The cognitive model of the monitoring-control process from the meta-meta layer to the meta layer.

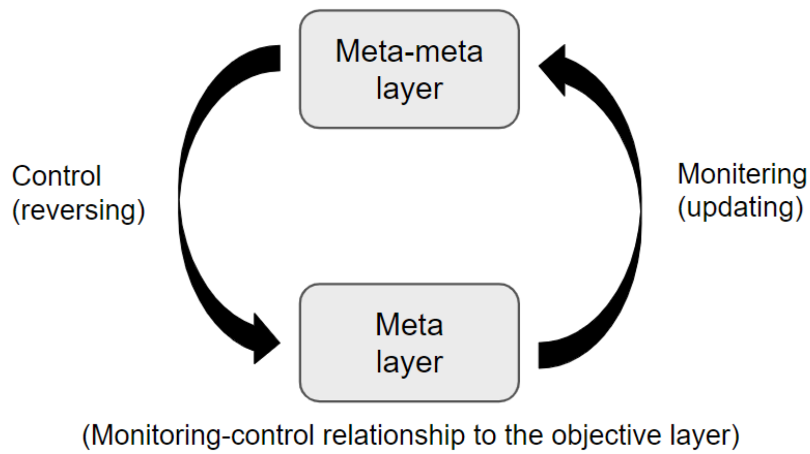


Figure 28. Illustration of the monitoring-control process at the meta-meta layer within Nelson and Naren's (1990) cognitive architecture.

The above architecture predicts a phenomenon regarding the tendency toward consistency between FOJs and the selection of cognitive offloading across SOJ levels. Specifically, the frequency of consistent trials (indicated by the consistency ratio) with higher SOJs is expected to exceed that of consistent trials with lower SOJs. This dissertation presents results that align with this assumption to some extent. Specifically, the consistency is lower than 0.5 for SOJ levels below 40, as indicated by Figure 24. This observation suggests that the SOJ at "40" might be a threshold for the meta-meta layer to regulate the meta layer, reversing the associations between FOJs and strategy choices for trials with SOJs rated from 0-30. However, this observation does not provide evidence to support the assumption fully, raising potential alternatives. One argument could point out the relatively lower consistency in SOJ levels of "90" (0.59) and "100" (0.52), showing a non-monotonic decrease from the highest to the lowest. Additionally, the consistency ratios at the boundary SOJ levels of 40 (0.52) and 30 (0.46) are too similar to indicate a clear threshold. These limitations could be attributed to two main factors. First, although SOJs can correct distorted FOJs, they may be susceptible to bias, which could unexpectedly affect the model (Frederick & Mochon, 2012). Second, the frequency of trials with SOJs ranging from 0-50 (relatively low SOJs) is considerably less than that of trials with SOJs ranging from 60-100 (relatively high SOJs). This unbalanced frequency between the relatively high and low SOJs might also weaken the strength of the observation in supporting the assumption. The SOJ elicitation process is expected to be improved to address these two concerns.

Individual participants accounted for substantial variability (57.97%) in strategy choices as a random effect. This observation suggests a critical role of individual differences in the associations between FOJs and strategy choices across various SOJ levels. Personal traits could underlie this individual difference by diversifying their propensity for employing cognitive offloading. Opposite traits within the same domain could cause some to be more conservative in offloading a learning target than others. For example, highly conscientious individuals may be more likely to retain information internally to maintain accuracy within the Big Five personality framework. In contrast, those who are less conscientious might rely more on external aid. These trait variations offer intriguing possibilities for further research.

While the interaction effect between FOJs and SOJs on strategy choices is significant, the interaction term may not substantially improve model fit, as indicated by Akaike's corrected information criterion and Bayesian information criterion values between Formulas 3 and 5 in Table 12. This observation may be due to the relatively small effect size of moderation by SOJs on the relationship between FOJs and strategy choices, suggesting that although the impact of FOJs on strategy choices depends on SOJs, this dependency does not add much explanatory power beyond the main effects of FOJs and SOJs alone. Such cases are common, where an interaction effect is statistically detectable but contributes little to overall model performance, particularly when the additive effects (including main effects and random effects) already explain most of the variance in the outcome. Specifically, in Formula 3, the additive effects alone explain 58.50% of the variance, indicating a high baseline R^2. This suggests that a substantial portion of the variation in strategy choices is already captured by these predictors. Thus, while the interaction term refines the relationship, it does not drive a substantial change in the outcome, making it theoretically meaningful but not practically necessary for prediction.

5.2 Alternatives

An alternative perspective to the current study is to consider whether the elicited SOJs indeed arise from higher-order metacognitive processes (at the meta-meta layer) or simply duplicate the FOJ elicitation at the meta layer using a different scale (e.g., as suggested by Zheng et al., 2023). Typically, a second-order judgement should rely on specific first-order information as input, provided explicitly from external sources. This would involve providing participants with explicit feedback on their performance during FOJ elicitation in the current research setting. However, the current study was limited in providing feedback during participation, as such information could alter the accuracy of metacognitive judgements in repeated trial-by-trial tasks. The absence of explicit first-order information might offer an alternative explanation for the observed FOJ-SOJ relationship: individuals' confidence levels could stem from their feelings about perceived memory performance rather than from a higher-order judgement facilitated by another functional component that incorporates new inputs external to the ongoing cognitive process. This dissertation assumes that if the SOJ merely duplicates the assessment of FOJ, its formation would involve translating the FOJ result onto the SOJ scale. This translation process leads to relatively consistent time durations for SOJ formation, which are unaffected by the complexity of the FOJ formation process. Moreover, given the substantial *SD* in FOJ formation times, there should be no significant correlation between FOJ and SOJ durations. SOJ formation would require a stable period dedicated solely to translation. In contrast, if FOJ and SOJ represent distinct judgements, the complex process underlying FOJ formation would logically require a similarly complex process for SOJ formation, leading to correlated time durations between these two sequential judgements. This latter relationship is supported by the findings from Study 1, indicating that the formations of the FOJ and SOJ are distinct processes. These results align with related research (Dunlosky et al., 2005; Recht et al., 2022).

An alternative approach to improve monitoring accuracy, rather than introducing SOJs to assess FOJs, is available. Research has shown that providing participants with feedback after test trials can effectively increase the accuracy of perceived memory performance in future learning trials – an approach known as metacognitive intervention (Carpenter et al., 2019). However, while metacognitive interventions improve monitoring accuracy, they may have a limited impact on the selection of cognitive offloading in those subsequent learning trials (Engeler & Gilbert, 2020). This dissertation suggests that individuals connect a present item with their previous experiences regarding similar items. In this case, previously corrected items from the intervention could provide the monitoring process with more accurate results. However, higher-order monitoring (e.g., at the meta-meta layer) should also seek evidence to indicate consistency between previous and current scenarios. Furthermore, accumulating this evidence may be challenging, as the two related scenarios could interact rather than remain independent. For example, even with identical paired associates, individuals might perceive changes in familiarity due to seeing the pair for the first versus the second time, leading them to view the two scenarios as clearly inconsistent. Consequently, the low consistency between scenarios before and after the intervention may cause confidence (even without an explicitly elicited SOJ) in continuing to monitor the association between perceived memory performance and strategy choice at a low level. This confidence could reverse that association, similar to how an SOJ would adjust an inaccurate FOJ. Therefore, the confidence level remains critical to regulating learning strategies, even when perceived memory performance is relatively accurate following the intervention.

5.3 Limitations

Specific methodological issues could compromise the results of metacognitive judgements and hinder the model's overall performance. First, while research related to cognitive offloading often presents metacognitive judgements without the cue component of a paired associate (Hu et al., 2019) or is inherently limited to providing cues during the elicitation of these judgements (Grinschgl et al., 2021b), most previous studies in the field of metacognition have elicited metacognitive judgements by prompting the cue (T. Nelson et al., 2004). Presenting the cue could be critical to the selection of cognitive offloading, as variations in FOJ accuracy – whether owing to the presence/absence of the cue or by the characteristics of the cue stimulus itself – may further influence strategy choices. An example of the current paradigm incorporating a cue during FOJ elicitation is shown in Figure 29. The present study also did not include cues during monitoring to increase the generalisability of the results across the cognitive offloading literature. Since the impact of cues on the selection of cognitive offloading remains unknown, this topic warrants further investigation.

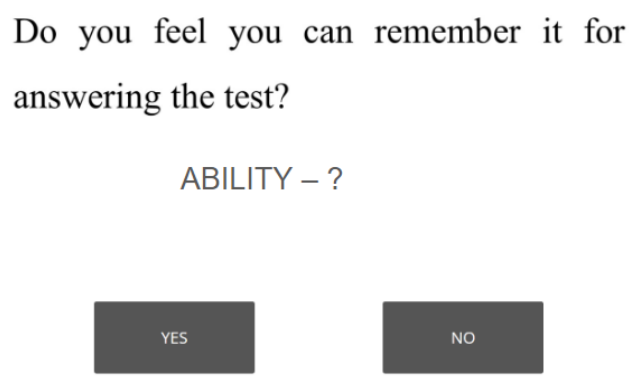


Figure 29. An example of the FOJ elicitation with a cue.

Second, the present study employed the paired-associate paradigm with a lower cognitive workload. While this study involved 48 pairs of English words as the learning material, related research in metacognitive monitoring typically involves materials with a heavier workload. For example, Dunlosky et al. (2005) included 66 pairs, Jersakova et al. (2017) included 90 pairs, and Hu et al. (2019) included 120 pairs. The reason for implementing a lighter version of the paired-associate tasks in this study was to avoid fatigue, as the impact of fatigue on the selection of cognitive offloading remains unclear. However, this lighter material could result in less challenging tasks for participants. Such less challenging tasks, with a reduced cognitive workload, might lead to different patterns of cognitive offloading compared with more demanding learning tasks. Future research could examine how fatigue and learning workload influence the selection of cognitive offloading.

Third, the present study focused on the role of the SOJ in the relationship between metacognitive judgements and the selection of cognitive offloading as exploratory research. This study employed a generalised linear mixed model to determine the mechanism for determining learning strategies on two layers of self-assessments, thereby addressing the issue of monitoring accuracy in the selection of cognitive offloading. However, this study did not find a significant correlation between the frequency of cognitive offloading and the time spent in the Retention session ($\rho = 0.17$, $p = 0.13$). This correlation could have implications, as it could support the argument that participants who employed less cognitive offloading during the Learning session might act more quickly in the subsequent sessions because of the pressure of forgetting. This insignificance of the correlation may be attributed to insufficient sample size, as the post hoc power analysis indicated that a sample size of 80 yields a statistical power of only 0.22 to detect such a weak correlation. Identifying this weak correlation with a larger sample size could provide insights into the monitoring-control relationship, warranting further research attention.

In addition, several limitations could be caused by online implementation. First, online platforms can make it challenging to ensure participants' return, complicating the management of remote participants. This challenge hinders the implementation of an effective within-subject design across sessions with long intervals (e.g., hours or days), preventing follow-up studies with the same participants from assessing memory performance without cognitive offloading. The difference in employing cognitive offloading, influenced by varying levels of actual memory performance, therefore remains unknown, impeding the ability to clarify the monitoring-control relationship. Additionally, the present research could not control the study process as rigorously as in a laboratory setting without face-to-face observation of the participants. For example, although platforms provide detection techniques and strict policies, it remains unfeasible to determine precisely whether participants actually wrote down certain items while reporting them as "Keeping in mind." This limitation potentially reduces the dataset reliability. Furthermore, the limitations of online platforms inhibit the effective use of dynamic rewards in the present study. In contrast, a laboratory environment remains the optimal choice for implementing such rewards in learning tasks to motivate participants to achieve better performance (e.g., participants who perform better in their learning tasks receive more financial rewards). Finally, the varying conditions between online and onsite participation may inherently influence the accuracy of FOJs (Cauvin et al., 2019), potentially leading to unknown effects on the selection of cognitive offloading. Future research should aim to replicate the current findings in a laboratory setting to address these concerns.

Finally, although the results in Study 3 were analysed on a trial-by-trial basis, they are still derived from a static group comparison (ANOVA) and a simple trend analysis (linear regression). While these analyses examine the stability of cognitive offloading behaviour across trials to detect broad trends, it assumes that each trial is independent of the previous one, limiting its ability to identify time-dependent effects (e.g., autocorrelation). This approach may fail to detect sudden changes in general preference for cognitive offloading at certain trials due to factors such as fatigue, experience, or serial position effects (e.g., primacy and recency effects). Additionally, the current approach assumes a uniform pattern across all participants, disregarding individual differences in the evolution of learning strategy patterns. Future research should further analyse trial-to-trial dependencies using time-series methods to gain a better understanding of the dynamic process of strategy choice.

Chapter 6 Conclusion

Overview

This chapter concludes the dissertation by addressing the following points: (1) a summary of current findings as the answer to the research question, along with perspectives from theoretical interpretation; (2) the academic contribution to knowledge science; (3) the impact on two related research areas in cognitive science concerning cognitive offloading; (4) a real-world educational scenario where cognitive offloading may offer practical implications; and (5) an outlook on future works, focusing on diversifying learning strategies to explore further the role of second-order metacognitive judgement (SOJ) within the monitoring-control architecture beyond cognitive offloading, potential methods for measuring confidence levels beyond introspective SOJ elicitation, as well as expanding the learning content to encompass a broader range of applied scenarios.

6.1 Summary

This dissertation presents findings describing the relationships between metacognitive factors in the selection of cognitive offloading as a learning strategy within the monitoring-control context, examining how individuals decide to employ cognitive offloading based on two layers of self-assessments. Two metacognitive factors – the perceived memory performance (signified by first-order metacognitive judgement, FOJ) and confidence level (signified by SOJ) – are proposed to play a role in this decision-making process. The key findings include the following: (1) perceived memory performance is negatively correlated with the selection of cognitive offloading, suggesting that individuals who feel their memory performance lower are more likely to offload; and (2) confidence level negatively moderates the relationship between perceived memory performance and the selection of cognitive offloading, suggesting that this connection is stronger on trials where individuals report higher SOJs.

Those findings conclude that, in learning tasks, individuals initially base their selection of cognitive offloading on their first-layer self-assessments (i.e., perceived memory performance) to some extent. However, the strength of their second-layer self-assessments (i.e., confidence in previous perceived memory performance on the same task) further regulates these initial strategy choices. This dissertation proposes that the significant interaction effect between FOJs and SOJs on strategy choices arises from the influence of confidence levels within the monitoring-control processes involved in SOJ formation, making the monitoring results more closely aligned with actual memory performance.

6.2 Contribution to knowledge science

Research on the cognitive psychology of metacognition has focused primarily on topics related to metacognitive strategies, addressing three key aspects: monitoring, control, and actual memory performance. Typical questions include "the processes that influence the accuracy of monitoring by various variables" (i.e., monitoring accuracy issues) and "the processes that influence control by the output of monitoring" (i.e., monitoring-control relationship issues), as reviewed by Norman et al. (2019). This principle also applies to discussing metacognition beyond the first layer (i.e., FOJ). From this perspective, related research has examined the influence of SOJs on FOJs to improve monitoring accuracy across various disciplines (Buratti & Allwood, 2015). Yet, there is no investigation into how SOJs impact control. The current research addresses this gap by demonstrating that SOJs influence the control process by regulating the strength of the association between FOJs and strategies, with cognitive offloading considered a conservative learning strategy. This dissertation deepens the understanding of the monitoring-control relationship by exploring the decision-making process for employing cognitive offloading through two layers of metacognitive judgements.

The current findings are closely related to concepts in knowledge science, with relevant definitions of knowledge clarified. Knowledge is commonly understood as rule-based thought oriented toward future tasks. This thought can take the form of propositional statements representing justified truths. A typical example of propositional knowledge in current research, represented by artificial truth, is the paired-associate format such as "ABILITY – CAPABILITY." Here, the paired associate represents the truth that a cue "ABILITY" is associated with a target "CAPABILITY." In this way, those who have learned the knowledge will recall “CAPABILITY” when presented with “ABILITY” during a learning task. The formation of propositional knowledge in the mind can thus be verified by observing performance on the learning task.

On another note, rule-based thought for performance in learning tasks can be procedural execution, which represents potential actions. Consider a scenario where individuals can access references to recall the targets of paired associates. In this case, if individuals know where and how to find the cue-target associations, their performance to recall the target may be equivalent to having remembered the paired associates precisely. Here, the paired associates serve as information, and the specific steps to access that information are represented in the mind as procedural knowledge, enabling effective retrieval actions. Procedural knowledge for accessing external information is a specific learning strategy performed as cognitive offloading.

Furthermore, individuals may sometimes have options for formatting knowledge with the same function but different representations. A typical example is the action for strategy choice in the current research. In the “Keeping in mind” case, individuals formatted the paired associates in their minds as propositional knowledge. In the “Writing down” case, they considered the paired associates as information on paper, with their awareness of the location where the target was written being regarded as procedural knowledge. While propositional knowledge provides the truth of the paired associate for later recall, procedural knowledge enables the steps needed to access the paired associate. Both types serve the same function during testing. The current work suggests that the decision for the selection between propositional and procedural knowledge is also informed by specific knowledge because it (1) is rule-based, (2) requires thought, and (3) is intended for future use. Significantly, this knowledge extends beyond the two types of knowledge related to learning strategies (at the object layer) and concerns individuals’ cognitive functioning (at the meta layer). Accordingly, this study regards knowledge at the meta layer as a new type of knowledge – *metaknowledge*.

Metaknowledge is crucial because its quality directly affects the formation of knowledge at the object layer, making it being both propositional or procedural. Research has consistently shown that perceived memory performance is inaccurate (Rhodes, 2016; Nelson & Dunlosky, 1991;

Tauber & Rhodes, 2012; Hu et al., 2019). However, a stable correlation appears between perceived memory performance and the selection of cognitive offloading (Risko & Dunn, 2015; Gilbert, 2015b; Boldt & Gilbert, 2019; Hu et al., 2019; Scott & Gilbert, 2024). Together, these findings suggest two potential consequences: (1) cognitive offloading may be less effective, and (2) the selection of cognitive offloading may be unpredictable to others. This can lead individuals to form knowledge at the object layer ineffectively. The current findings regarding two-layer metacognitive judgements within the monitoring-control process may help address the quality issue of metaknowledge and, in turn, improve the processes of knowledge creation at the object layer.

These findings contribute to knowledge science. Knowledge science systematically studies knowledge, focusing on its creation, accumulation, and utilisation. Various theories within this discipline have explored knowledge creation processes across fundamental domains (Nonaka & Takeuchi, 1995; Nonaka et al., 2014; Young, 2012; Maruta, 2014). To date, three primary knowledge fields have been proposed to classify individual-level knowledge during the creation process, namely, cognitive (knowledge regarding what is), emotional (knowledge regarding how we feel), and spiritual (knowledge regarding aspirations), as outlined by the triple helix knowledge perspective (Bratianu, 2013; Bratianu & Orzea, 2014; Bratianu, 2015). The current research expands upon the triple knowledge fields by introducing an additional type known as metaknowledge, along with an approach for enhancing its quality. This extension provides a more comprehensive understanding of knowledge fields, deepening the insights into the dynamic interconnections within individual-level knowledge creation processes.

6.3 Academic implications

The current findings may impact two research areas concerning cognitive offloading: (1) the causal relationship between metacognitive judgements and the selection of cognitive offloading and (2) the optimisation of employing cognitive offloading.

While previous research has shown that manipulation of monitoring can influence learning behaviours for specific strategies (Metcalfe, 2009), similar connections between perceived memory performance and the selection of cognitive offloading have not been established (Grinschgl et al., 2021b). This absence of findings suggests that metacognitive judgements may be epiphenomenal, reflecting feelings beyond learning behaviours rather than determinants influencing these behaviours. Verifying a causal monitoring-control relationship in the context of cognitive offloading is, therefore, essential. The current work contributes to related research by introducing the concept of the SOJ, suggesting that operations affecting FOJs in related studies may not similarly influence SOJs. This dissociation between FOJs and SOJs could hinder systematic changes in strategy choices, as SOJs may further regulate those experimentally manipulated FOJs in unexpected ways.

Related research has revealed that individuals tend to employ cognitive offloading more frequently than necessary rather than optimising their strategy choices according to a cost-benefit trade-off within the monitoring-control framework (Gilbert et al., 2020). The current work contributes to this topic by introducing the concept of SOJ, suggesting that SOJs may interact with FOJs to guide strategy choices based on factors beyond the traditional cost-benefit frameworks used in experimental settings. Since SOJs monitor metacognition, biassed strategy choices may be attributed to an imbalance in individuals' understanding of mental versus environmental contexts. Specifically, individuals may better understand their environments and tools than their own cognitive and metacognitive processes. This assumption suggests that research exploring the optimised use of cognitive offloading should consider factors beyond cost-benefit analysis within the framework of human rationality.

6.4 Practical implications

The findings from this research deepen the understanding of the mechanisms underlying human memory, particularly the interaction between FOJs and SOJs in the selection of cognitive offloading. While theoretical in nature, this study has noteworthy implications for educational practice, especially in designing educational technologies intended to help learners manage cognitive load more effectively. In technology-assisted learning environments, such as interactive apps, learning management systems, or smart tutors powered by artificial intelligence algorithms, learners frequently engage in cognitive offloading – using tools or external resources to reduce the burden on their memory (e.g., working memory). This research suggests that learners' decisions to employ cognitive offloading (e.g., consulting external aids such as calculators, apps, or digital notes) are influenced not only by their FOJs that reflect their immediate sense of cognitive capacity, but also by their SOJs that monitor and control the appropriateness of these decisions. A key implication for technology design is the limitation of tools that rely solely on learners' FOJs.

Specifically, technology-assisted learning tools could be intuitively designed to assess learners' perceived performance through self-reported feedback and, in turn, adapt to their cognitive state based on what they reported. For example, students with different perceived memory performances may have varying attitudes toward using systems to store or organise information, such as saving key points, setting reminders, or creating visual aids like concept maps. Those who feel they can handle the cognitive load might be more independent of such aids, whereas those who believe less might rely more on the system to offload memory tasks.

However, simply prompting learners to self-report their perceived performance may not be sufficient, as learners may not always accurately judge their cognitive ability (i.e., at the first-order layer). In such cases, even when learners report that they can remember the target, which suggests unnecessary cognitive offloading options, they may still expect to offload tasks. Without an effective SOJ to detect such expectations accurately, the system may interpret inconsistencies between self-reports and actual learning preferences as uncontrollable factors, similar, for example, to an overreliance on cognitive offloading resulting from individual differences. This could hinder the system's development as an intelligent, intuitive, and user-friendly tool for high user experience. Incorporating the SOJ approach into a tailored feedback loop could help understand learners' preferences for offloading in this context.

6.5 Future suggestions

This dissertation focused on two layers of self-assessment, represented by FOJs and SOJs, within the monitoring-control relationship, considering cognitive offloading as the conservative option in employing learning strategies. Future research should further explore the monitoring-control relationship through high-order metacognitive judgements, incorporating a broader array of conservative learning strategies for control, such as material review, time allocation, and goal setting (Kornell & Bjork, 2007). Although conservative learning strategies beyond cognitive offloading are outside the scope of this dissertation, further exploration of these strategies could shed light on the monitoring-control relationship. For example, future work could replicate the current paradigm by substituting cognitive offloading with material review as the conservative learning strategy. The research process could remain largely the same in the current approach, with the strategy options adjusted to “Review once more” and “Proceed without review.” Here, the conservative option, “Review once more,” would involve reviewing the paired associate once more before the Test session, whereas the standard option, “Proceed without review,” would advance directly to the subsequent trial, completing the current learning task. For an example created by Gorilla, see Figure 30. In this way, the FOJ-SOJ relationship in the selection of a material review can be systematically investigated. Similarly, other learning strategies could be explored following this approach. This dissertation anticipates that the differences identified among these strategies will yield more profound insights into the monitoring-control relationship.

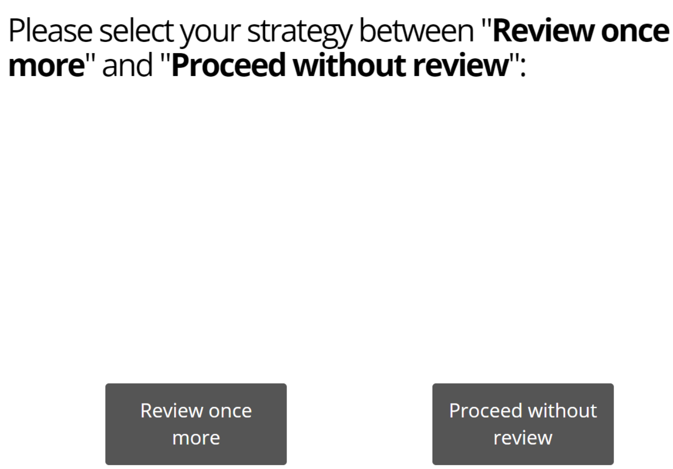


Figure 30. The strategy choice includes options “Review once more” and “Proceed without review.”

The elicitation of confidence levels can extend beyond introspective reports. Various methods beyond scales could be explored in future research. Examples include structured interviews, peer evaluations, decision times, eye-tracking data (e.g., gaze tracking or pupil dilation), neural feedback (e.g., electroencephalography or functional near-infrared spectroscopy), and gesture-based reactions. These methods may offer advantages depending on the practical setting. While the current research focuses primarily on metacognition through introspective reports within the paired-associate paradigm, exploring diverse methods for eliciting confidence could expand approaches for metacognitive judgements. This is especially relevant for applied settings where scales may not be optimal. For example, eye-tracking data could serve as an effective alternative for assessing confidence by monitoring gaze patterns and pupil dilation during the formation of FOJs. This could be implemented by utilising a computer equipped with a table-mounted eye-tracking system that captures participants' eye movements on the monitor. Refer to Figure 31 for an illustrative example of the device and its operating environment. In this approach, gaze fixation duration on specific areas of interest could indicate varying confidence levels; i.e., longer fixation durations might suggest uncertainty. Pupil dilation could also serve as a physiological marker, with, say, larger dilations that might be associated with increased uncertainty. Furthermore, eye-tracking systems could enable real-time confidence detection, providing a clearer understanding of the processes underlying metacognitive judgements with higher temporal resolution. In addition to eye tracking, other methods may provide valuable insights into confidence across diverse practical settings, potentially surpassing self-report scales in terms of accessibility. Future research could evaluate the effectiveness of these methods in assessing confidence, contributing to enhanced methodological practices.

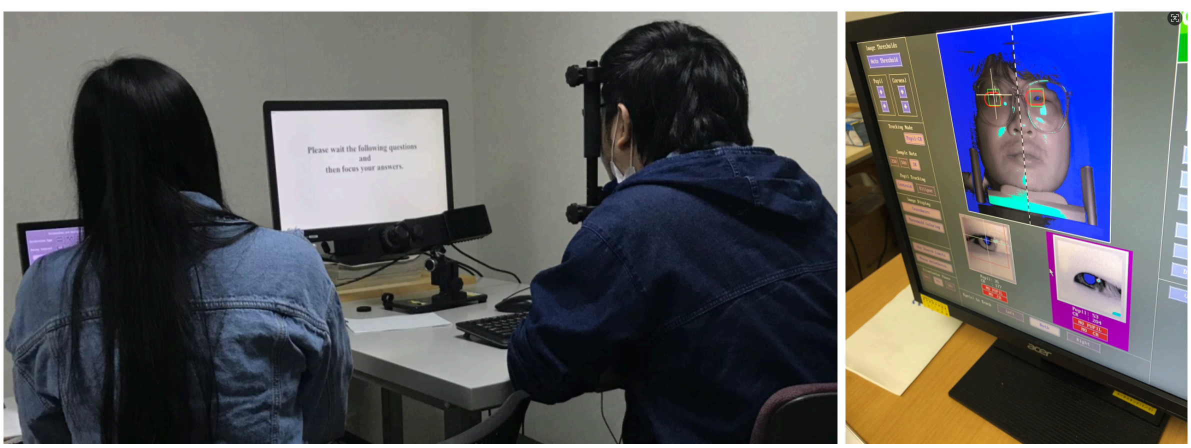


Figure 31. The example implementation involves conducting learning tasks using a high-performance table-mounted eye-tracking system (EyeLink 1000 Plus) in a sound-proof room.

The present study employed the paired-associate paradigm for learning tasks, using English word pairs as learning materials. While this paradigm is valuable in memory and metacognition research (Nelson & Dunlosky, 1991; Dunlosky et al., 2025; Jersakova et al., 2017; Yang et al., 2018; Senkova & Otani, 2021), its content is often far removed from practical applications. Instead, various alternatives could better align with real-world contexts while still engaging associative memory mechanisms. For instance, in second language learning, native words could serve as cues and foreign words as targets (e.g., Hello – Konnichiwa for Japanese learners who speak English, for example created by Gorilla, see Figure 32). Similarly, the paradigm could be adapted for STEM (Science, Technology, Engineering, and Mathematics) education by pairing scientific terms with their definitions (e.g., Mitochondria – Powerhouse of the cell), chemical elements with their symbols (e.g., Gold – Au), or mathematical formulas with their applications (e.g., Pythagorean theorem – $a^2 + b^2 = c^2$). Beyond education, alternative content could extend to everyday contexts, such as associating people's details with their names (e.g., Neighbor with the red car – John) or using photos of faces as cues for name recall (i.e., the *face-name association*, Ma et al., 2019). Future research should explore these alternatives to enhance the practical implications of the paired-associate paradigm.

Figure 32A. The paired associate for Japanese learners who speak English in the Learning session.

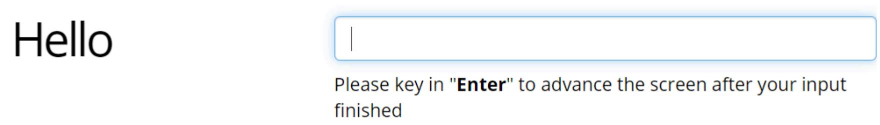


Figure 32B. The paired associate for Japanese learners who speak English in the Test session.

Acknowledgements

My time at JAIST has been a profoundly transformative chapter. I could not have achieved this without the generous support of various influences.

I would like to begin by sincerely thanking my supervisor, Prof Dr Tsutomu Fujinami, whose mentorship has been the cornerstone of this work. I vividly recall our first email exchange in the fall of 2017 while I was in Beijing. Since then, Sensei has cultivated in me the core principles essential for becoming a capable scientist and scholar. I am deeply fortunate to have had a dedicated mentor throughout this path.

I sincerely thank my committee members, Prof Dr Dam Hieu Chi, Prof Dr Mizumoto Masaharu, and Prof Dr Yoshioka Hidekazu, and Prof Dr Kojima Haruyuki from Kanazawa University, for their thorough review, which greatly strengthened this dissertation. I appreciate my second supervisor, Prof Dr Shohei Hidaka, for his insightful feedback on my research proposal, which inspired me with fresh ideas. I also appreciate my minor project supervisor, Prof Dr Hideaki Kanai, whose guidance helped me shift my perspective from psychology to engineering.

I am grateful to Dr Akemi Tera, a researcher at Fujinami Lab, whose enlightening discussions on eye-tracking over the years have enriched my research. I also thank my lab mates for their collaboration and my colleagues at the ISC Center for their camaraderie during my time there from 2020 to 2023. Additionally, I appreciate the financial support from JAIST, including the JAIST Research Grant (Fundamental Research) 2020, the JAIST Research Grant (Fundamental Research) 2021, as well as the Doctoral Research Fellowship (2020-2023).

I am also grateful to Mr Jeffrey Keith Spaneas Bland, CTO at FOVE Inc., for sharing insights from both engineering and management perspectives, which helped keep each research project running smoothly. I equally thank Mr Yoichiro Nishina, COO at FOVE Inc., for inspiring me to consider the practical implications of research in greater depth.

I am deeply thankful to my family for always standing by me through every challenge. I especially want to express my gratitude to my grandmother, whose unwavering encouragement has been a constant source of strength. To my mother, thank you for providing the necessary support whenever my concerns caused hesitation, always helping me move forward. Lastly, I am forever thankful to my wife, whose care and love have accompanied me every step of the way.

A special mention to my cat, whose mere presence recharges me, even when just in my thoughts. I am thankful for the joy he brings into my life.

My Catholic faith has shaped my approach throughout this path, for which I am truly grateful.

Publication List

Journals

[1]. ○Ma, Y., & Fujinami, T. (2024). The influence of second-order metacognitive judgments on cognitive offloading within the monitoring-control relationship. Discover Psychology, 4(1), 136. https://doi.org/10.1007/s44202-024-00258-8

[2]. ○Ma, Y., Bland, J. K. S., & Fujinami, T. (2024). Classification of Alzheimer's Disease and Frontotemporal Dementia Using Electroencephalography to Quantify Communication between Electrode Pairs. Diagnostics, 14(19), 2189.
https://doi.org/10.3390/diagnostics14192189

[3]. Mizukami, K., Taguchi, M., Kouketsu, T., Sato, N., Tanaka, Y., Iwakiri, M., Nishina, Y., ○Ma, Y., Chernyak, I., & Karaki, S. (2024). A Cognitive Function Test Utilizing Eye-tracking Technology in Virtual Reality is useful to distinguish between normal cognition, MCI and mild dementia. Archives of Gerontology and Geriatrics Plus, 100070.
https://doi.org/10.1016/j.aggp.2024.100070

International conferences

[1]. ○Ma, Y., Bland, J., Yoshikawa, G, and Fujinami, T. Quantifying Consciousness for Alzheimer's Disease Diagnosis through Electroencephalogram Processing. 2024 8th International Conference on Medical and Health Informatics, May 2024, Yokohama, Japan. (peer reviewed, oral)
https://doi.org/10.1145/3673971.3673978

[2]. ○Ma, Y., Yoshikawa, G., and Bland, J. K. S. The Potential of Measuring Conscious States via EEG Closed Eyes Recordings as A Diagnostic Marker of Alzheimer's Disease. International Conference on Alzheimer's and Parkinson's Diseases and Related Neurological Disorders, March 2024, Lisbon, Portugal. (peer reviewed, virtual poster)

[3]. Miyakawa, T., Bland, J., Yoshikawa, G., and ○Ma, Y. A Rapid VR House Interior Generator with Eye-tracking Feedback for Home Staging. International Conference on Artificial Reality

and Telexistence Eurographics Symposium on Virtual Environments (2023), Nov 2023, Dublin, Ireland. (peer reviewed, poster, interaction)
https://doi.org/10.2312/egve.20231330
[4]. ○Ma, Y., Hu, N., Kanai, H., and Fujinami, T. Reliabilities of Eye Fixations for Assessing Short-term Memory on Forward Span Tasks. 2023 7th International Conference on Medical and Health Informatics. May 2023, Kyoto, Japan. (peer reviewed, oral)
https://doi.org/10.1145/3608298.3608303
[5]. ○Ma, Y., and Fujinami, T. Visual fixation-based eye-tracking features in memory-updating tasks for assessing short-term memory. 11th Conference of the International Society for Affective Disorders, November 2021, online. (peer reviewed, virtual poster)
[6]. ○Ma, Y., Zhang, X., and Fujinami, T. Measuring working memory by associating implicit learning process to memory-update tasks. 21st WPA World Congress of Psychiatry, October 2021, Geneva, Switzerland. (peer reviewed, virtual poster)
[7]. Bao, H., Yabuuchi, K., ○Ma, Y., and Nagai, Y. A qualitative study on the style of art teaching instruction for improving creativity: the perspectives from Chinese elementary school. The 11th Asian Conference on Psychology & the Behavioral Sciences, March 2021, Tokyo, Japan. (peer reviewed, oral pre-recorded)
[8]. ○Ma, Y., Zhang, X., and Fujinami, T. A modified n-back task for measuring working memory via eye-tracker. 20th WPA World Congress of Psychiatry, March 2021, Bangkok, Thailand. (peer reviewed, virtual poster)
[9]. ○Ma, Y., Zhang, X., and Fujinami, T. A comparive study of self-reporting methods for assessing mind-wandering state. 15th International Conference on Knowledge, Information and Creativity Support Systems, November 2020, Hobart, Australia. (peer reviewed, oral live stream)
[10]. ○Ma, Y., and Fujinami, T. The framing effect during dynamic decision-making process. The 10th Asian Conference on Psychology & the Behavioral Sciences, March 2020, Tokyo, Japan. (peer reviewed, oral pre-recorded)
[11]. ○Ma, Y., Wang, Q., and Fujinami, T. A preliminary study of cognitive training by error-less learning in virtual reality. 2020—2nd International Conference on Research in Life-Sciences & Healthcare, March 2020, Melbourne, Australia. (non-peer reviewed, oral pre-recorded)

[12]. ○Ma, Y., and Fujinami, T. A new paradigm of N back task for the focus of attention. International Conference on Emotion and Sensibility 2019, November 2019, Nagano, Japan. (peer reviewed, poster)

Domestic conferences

[1]. Miyakawa, T., and ○Ma, Y. Combining 3D Models and Panoramic Images in VR Spaces Gathering of Line-of-sight Data and Visualization (Non-peer reviewed, Oral & Poster). The 46th Symposium on Computer Technology of Information, Systems and Applications, December 2023, Tokyo. (in Japanese)
[2]. Yan, F., ○Ma, Y., and Fujinami, フィッシングメールが人を欺く要因. 第46回セキュリティ心理学とトラスト研究発表会, March 2022, 那覇. (non-peer reviewed, oral) (in Japanese) http://id.nii.ac.jp/1001/00216542/
[3]. Yan, F., ○Ma, Y., and Fujinami, T.フィッシングメールに対する危険判断と言葉の感性に関する考察. 第67回ことば工学研究会, December 2021, Tokyo. (non-peer reviewed, oral) (in Japanese)
[4]. ○Ma, Y., Zhang, X., and Fujinami, T. 認知症の症状を積極的に緩和するVRゲーム. Matching HUB Kanazawa 2020, November 2020, Kanazawa. (non-peer reviewed, poster) (in Japanese)
[5]. ○Ma, Y., Lai, K., and Fujinami, T. Public figures as training material for error-less learning for MCI. SIG SKL-27, March 2019, Kanazawa. (non-peer reviewed, oral)

Awards

[1]. Granted Japanese-Language Proficiency Test (JLPT) Grant by JAIST in 2022
[2]. Granted JAIST Research Grants (Fundamental Research) 2021 by JAIST from August 2021 to March 2022 (JPY 250,000)
[3]. Awarded Kunifuji Award 2020 by KICSS 2020 in November 2020
[4]. Granted JAIST Research Grants (Fundamental Research) 2020 by JAIST from August 2020 to March 2021 (JPY 500,000)

[5]. Granted Doctoral Research Fellow by JAIST from April 2020 to March 2023 (JPY 890,000 per year)

[6]. Granted JAIST Research Grants (Travel Grant) by JAIST in 2020

[7]. Awarded ICES 2019 Encouragement Poster Presentation Award by ICES 2019 in November 2019

Ethics Declarations

The study was conducted in accordance with the Declaration of Helsinki and approved by the Life Science Committee of Japan Advanced Institute of Science and Technology (approval granted on May 31, 2022, under protocol 人04-007, conducted from June 15, 2022, to March 31, 2023). Informed consent was obtained from all the subjects involved in the study.

References

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

Arango-Muñoz, S. (2013). Scaffolded memory and metacognitive feelings. *Review of Philosophy and Psychology*, *4*, 135-152. https://doi.org/10.1007/s13164-012-0124-1

Arbuckle, T. Y., & Cuddy, L. L. (1969). Discrimination of item strength at time of presentation. *Journal of experimental psychology*, *81*(1), 126. https://doi.org/10.1037/h0027455

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In *Psychology of learning and motivation* (Vol. 2, pp. 89-195). Academic press.

Baddeley, A. (1992). Working memory. *Science*, *255*(5044), 556-559.

Ballard, D. H., Hayhoe, M. M., & Pelz, J. B. (1995). Memory representations in natural tasks. *Journal of cognitive neuroscience*, *7*(1), 66-80. https://doi.org/10.1162/jocn.1995.7.1.66

Bliss, J. C., Crane, H. D., Mansfield, P. K., & Townsend, J. T. (1966). Information available in brief tactile presentations. *Perception & Psychophysics*, *1*(4), 273-283. https://doi.org/10.3758/BF03207391

Boldt, A., & Gilbert, S. J. (2019). Confidence guides spontaneous cognitive offloading. *Cognitive Research: Principles and Implications*, *4*(1), 1-16. https://doi.org/10.1186/s41235-019-0195-y

Bratianu, C. (2013). The triple helix of the organizational knowledge. *Management dynamics in the knowledge economy*, *1*(2), 207-207.

Bratianu, C., & Orzea, I. (2014). Emotional knowledge: The hidden part of the knowledge iceberg. *Management Dynamics in the Knowledge Economy*, *2*(1), 41-56.

Bratianu, C. (2015). *Organizational Knowledge Dynamics: Managing Knowledge Creation, Acquisition, Sharing, and Transformation*. Hershey: IGI Global.

Brown, A. L. (1978). *Knowing when, where, and how to remember; a problem of metacognition*. Advances in instructional psychology.

Brown, A. L. (1980). Metacognitive development and reading. Theoretical issues in reading comprehension: *Perspectives from cognitive psychology, linguistics, artificial intelligence, and education*, *12*, 453-481.

Buratti, S., & Allwood, C. M. (2012). The accuracy of meta-metacognitive judgments: Regulating the realism of confidence. *Cognitive processing*, *13*, 243-253. https://doi.org/10.1007/s10339-012-0440-5

Buratti, S., Allwood, C. M., & Kleitman, S. (2013). First-and second-order metacognitive judgments of semantic memory reports: The influence of personality traits and cognitive styles. *Metacognition and learning*, *8*, 79-102. https://doi.org/10.1007/s11409-013-9096-5

Buratti, S., & Allwood, C. M. (2015). Regulating metacognitive processes—support for a meta-metacognitive ability. *Metacognition: Fundaments, Applications, and Trends: A profile of the current State-of-the-Art*, 17-38. https://doi.org/10.1007/978-3-319-11062-2_2

Carpenter, J., Sherman, M. T., Kievit, R. A., Seth, A. K., Lau, H., & Fleming, S. M. (2019). Domain-general enhancements of metacognitive ability through adaptive training. *Journal of Experimental Psychology: General*, *148*(1), 51. https://doi.org/10.1037/xge0000505

Cauvin, S., Moulin, C., Souchay, C., Schnitzspahn, K., & Kliegel, M. (2019). Laboratory vs. naturalistic prospective memory task predictions: young adults are overconfident outside of the laboratory. *Memory*, *27*(5), 592-602. https://doi.org/10.1080/09658211.2018.1540703

Connerton, P. (2008). Seven types of forgetting. *Memory studies*, *1*(1), 59-71.

Cowan, N., Elliott, E. M., Saults, J. S., Morey, C. C., Mattox, S., Hismjatullina, A., & Conway, A. R. (2005). On the capacity of attention: Its estimation and its role in working memory and cognitive aptitudes. *Cognitive psychology*, *51*(1), 42-100. https://doi.org/10.1016/j.cogpsych.2004.12.001

Crawford, J. D., & Stankov, L. (1996). Age differences in the realism of confidence judgements: A calibration study using tests of fluid and crystallized intelligence. *Learning and*

Individual Differences, *8*(2), 83-103.
https://doi.org/10.1016/S1041-6080(96)90027-8

Darwin, C. J., Turvey, M. T., & Crowder, R. G. (1972). An auditory analogue of the Sperling partial report procedure: Evidence for brief auditory storage. *Cognitive Psychology*, *3*(2), 255-267.
https://doi.org/10.1016/0010-0285(72)90007-2

Double, K. S., Birney, D. P., & Walker, S. A. (2018). A meta-analysis and systematic review of reactivity to judgements of learning. *Memory*, *26*(6), 741-750.
https://doi.org/10.1080/09658211.2017.1404111

Dunlosky, J., Serra, M. J., Matvey, G., & Rawson, K. A. (2005). Second-order judgments about judgments of learning. *The Journal of general psychology*, *132*(4), 335-346.
https://doi.org/10.3200/GENP.132.4.335-346

Dunlosky, J., & Connor, L. T. (1997). Age differences in the allocation of study time account for age differences in memory performance. *Memory & cognition*, *25*, 691-700.
https://doi.org/10.3758/BF03211311

Dunn, T. L., & Risko, E. F. (2016). Toward a metacognitive account of cognitive offloading. *Cognitive Science*, *40*(5), 1080-1127.
https://doi.org/10.1111/cogs.12273

Engeler, N. C., & Gilbert, S. J. (2020). The effect of metacognitive training on confidence and strategic reminder setting. *PLoS One*, *15*(10), e0240858.
https://doi.org/10.1371/journal.pone.0240858

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. *Behavior research methods*, *41*(4), 1149-1160.
https://doi.org/10.3758/BRM.41.4.1149

Fiedler, K., Ackerman, R., & Scarampi, C. (2019). Metacognition: Monitoring and controlling one's own knowledge, reasoning and decisions. *The psychology of human thought: An introduction*, 89-111.
https://doi.org/10.17885/heiup.470.c6669

Flavell, J. H. (1976). Metacognitive aspects of problem solving. *The nature of intelligence*.

Flavell, J. H. (1978). Metacognitive development. *Structural/process theories of complex human* *behavior*, 213-245.

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American psychologist*, *34*(10), 906. https://doi.org/10.1037/0003-066X.34.10.906

Frederick, S. W., & Mochon, D. (2012). A scale distortion theory of anchoring. *Journal of Experimental Psychology: General*, *141*(1), 124. https://doi.org/10.1037/a0024006

Fritzsche, E. S., Händel, M., & Kröner, S. (2018). What do second-order judgments tell us about low-performing students' metacognitive awareness?. *Metacognition and Learning*, *13*, 159-177. https://doi.org/10.1007/s11409-018-9182-9

Fujinami, T., & Hidaka, S. (2019). A representation of rhythmic motions. *New Generation Computing*, *37*, 185-201. https://doi.org/10.1007/s00354-019-00051-5

Gilbert, S. J. (2015a). Strategic offloading of delayed intentions into the external environment. *Quarterly journal of experimental psychology*, *68*(5), 971-992. https://doi.org/10.1080/17470218.2014.972963

Gilbert, S. J. (2015b). Strategic use of reminders: Influence of both domain-general and task-specific metacognitive confidence, independent of objective memory ability. *Consciousness and Cognition*, *33*, 245-260. https://doi.org/10.1016/j.concog.2015.01.006

Gilbert, S. J., Boldt, A., Sachdeva, C., Scarampi, C., & Tsai, P. C. (2023). Outsourcing memory to external tools: A review of 'intention offloading'. *Psychonomic Bulletin & Review*, *30*(1), 60-76. https://doi.org/10.3758/s13423-022-02139-4

Gilbert, S. J., Bird, A., Carpenter, J. M., Fleming, S. M., Sachdeva, C., & Tsai, P. C. (2020). Optimal use of reminders: Metacognition, effort, and cognitive offloading. *Journal of Experimental Psychology: General*, *149*(3), 501. https://doi.org/10.1037/xge0000652

Grinschgl, S. (2020). *Determinants and consequences of offloading working memory processes* (Doctoral dissertation, Universität Tübingen).

Grinschgl, S., Papenmeier, F., & Meyerhoff, H. S. (2021a). Consequences of cognitive offloading: Boosting performance but diminishing memory. *Quarterly Journal of Experimental Psychology*, *74*(9), 1477-1496. https://doi.org/10.1177/17470218211008060

Grinschgl, S., Meyerhoff, H. S., Schwan, S., & Papenmeier, F. (2021b). From metacognitive beliefs to strategy selection: does fake performance feedback influence cognitive offloading?. *Psychological Research*, *85*, 2654-2666. https://doi.org/10.1007/s00426-020-01435-9

Händel, M., & Fritzsche, E. S. (2016). Unskilled but subjectively aware: Metacognitive monitoring ability and respective awareness in low-performing students. *Memory & Cognition*, *44*, 229-241. https://doi.org/10.3758/s13421-015-0552-0

Hart, J. T. (1965). Memory and the feeling-of-knowing experience. *Journal of educational psychology*, *56*(4), 208. https://doi.org/10.1037/h0022263

Hart, J. T. (1967). Memory and the memory-monitoring process. *Journal of verbal learning and verbal behavior*, *6*(5), 685-691. https://doi.org/10.1016/S0022-5371(67)80072-0

Hines, J. C., Touron, D. R., & Hertzog, C. (2009). Metacognitive influences on study time allocation in an associative recognition task: An analysis of adult age differences. *Psychology and Aging*, *24*(2), 462.

Hu, X., Luo, L., & Fleming, S. M. (2019). A role for metamemory in cognitive offloading. *Cognition*, *193*, Article 104012. https://doi.org/10.1016/j.cognition.2019.104012

Jersakova, R., Allen, R. J., Booth, J., Souchay, C., & O'Connor, A. R. (2017). Understanding metacognitive confidence: Insights from judgment-of-learning justifications. *Journal of Memory and Language*, *97*, 187-207. https://doi.org/10.1016/j.jml.2017.08.002

Keith, M. G., Tay, L., & Harms, P. D. (2017). Systems perspective of Amazon Mechanical Turk for organizational research: Review and recommendations. *Frontiers in psychology*, *8*, 1359.

https://doi.org/10.3389/fpsyg.2017.01359

Klein, S. B. (2018). *Learning: Principles and applications*. Sage Publications.

Kornell, N., & Bjork, R. A. (2007). The promise and perils of self-regulated study. *Psychonomic bulletin & review*, *14*(2), 219-224. https://doi.org/10.3758/BF03194055

Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of experimental psychology: General*, *126*(4), 349. https://doi.org/10.1037/0096-3445.126.4.349

Ma, Y., & Fujinami, T. (2024). The influence of second-order metacognitive judgments on cognitive offloading within the monitoring-control relationship. *Discover Psychology*, *4*(1), 136. https://doi.org/10.1007/s44202-024-00258-8

Ma, Y., Hu, N., Kanai, H., & Fujinami, T. (2023, May). Reliabilities of Eye Fixations for Assessing Short-term Memory on Forward Span Tasks. In *Proceedings of the 2023 7th International Conference on Medical and Health Informatics* (pp. 25-28). https://doi.org/10.1145/3608298.3608303

Ma, Y., Lai, K., & Fujinami, T. (2019). Public figures as training material for error-less learning for MCI (Vol. 27, pp. 1-5). SIG-SKL.

Malmberg, K. J., Raaijmakers, J. G., & Shiffrin, R. M. (2019). 50 years of research sparked by Atkinson and Shiffrin (1968). *Memory & cognition*, *47*, 561-574. https://doi.org/10.3758/s13421-019-00896-7

Maruta, R. (2014). The creation and management of organizational knowledge. *Knowledge-Based Systems*, *67*, 26-34. https://doi.org/10.1016/j.knosys.2014.06.012

Mathy, F., Chekaf, M., & Cowan, N. (2018). Simple and complex working memory tasks allow similar benefits of information compression. *Journal of Cognition*, *1*(1). https://doi.org/10.5334/joc.31

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

McGaugh, J. L. (2000). Memory--a century of consolidation. *Science*, *287*(5451), 248-251.

Melton, A. W. (1963). Implications of short-term memory for a general theory of memory. *Journal of verbal Learning and verbal Behavior*, *2*(1), 1-21. https://doi.org/10.1016/S0022-5371(63)80063-8

Metcalfe, J. (2009). Metacognitive judgments and control of study. *Current directions in psychological science*, *18*(3), 159-163. https://doi.org/10.1111/j.1467-8721.2009.01628.x

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

Miller, T. M., & Geraci, L. (2011). Unskilled but aware: reinterpreting overconfidence in low-performing students. *Journal of experimental psychology: learning, memory, and cognition*, *37*(2), 502. https://doi.org/10.1037/a0021802

Mochon, D., & Frederick, S. (2013). Anchoring in sequential judgments. *Organizational Behavior and Human Decision Processes*, *122*(1), 69-79. https://doi.org/10.1016/j.obhdp.2013.04.002

Nederhand, M. L., Tabbers, H. K., De Bruin, A. B., & Rikers, R. M. (2021). Metacognitive awareness as measured by second-order judgements among university and secondary school students. *Metacognition and Learning*, *16*, 1-14. https://doi.org/10.1007/s11409-020-09228-6

Nelson, T. O. (1996). Consciousness and metacognition. *American psychologist*, *51*(2), 102. https://doi.org/10.1037/0003-066X.51.2.102

Nelson, T. O., & Dunlosky, J. (1991). When people's judgments of learning (JOLs) are extremely accurate at predicting subsequent recall: The “delayed-JOL effect”. *Psychological Science*, *2*(4), 267-271. https://doi.org/10.1111/j.1467-9280.1991.tb0014

Nelson, T. O. & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In *Psychology of learning and motivation (Vol. 26, pp. 125-173)*. Academic Press.

Nelson, T. O., & Narens, L. (1994). Why investigate metacognition. *Metacognition: Knowing about knowing*, *13*, 1-25.

Nelson, T. O., Narens, L., & Dunlosky, J. (2004). A revised methodology for research on metamemory: Pre-judgment Recall and Monitoring (PRAM). *Psychological methods, 9*(1), 53.

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1998). The University of South Florida word association, rhyme, and word fragment norms. http://w3.usf.edu/FreeAssociation/Intro.html

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

Nonaka, I., & Takeuchi, H. (1995). *The knowledge-creating company: How Japanese companies create the dynamics of innovation*. Oxford university press.

Nonaka, I., Kodama, M., Hirose, A., & Kohlbacher, F. (2014). Dynamic fractal organizations for promoting knowledge-based transformation–A new paradigm for organizational theory. *European Management Journal*, *32*(1), 137-146. https://doi.org/10.1016/j.emj.2013.02.003

Norman, E., Pfuhl, G., Sæle, R. G., Svartdal, F., Låg, T., & Dahl, T. I. (2019). *Metacognition in psychology*. Review of General Psychology, *23*(4), 403-424. https://doi.org/10.1177/1089268019883

Peer, E., Vosgerau, J., & Acquisti, A. (2014). Reputation as a sufficient condition for data quality on Amazon Mechanical Turk. *Behavior research methods*, *46*, 1023-1031. https://doi.org/10.3758/s13428-013-0434-y

Peer, E., Rothschild, D., Gordon, A., Evernden, Z., & Damer, E. (2022). Data quality of platforms and panels for online behavioral research. *Behavior Research Methods*, *1*. https://doi.org/10.3758/s13428-021-01694-3

Prisko, L. H. (1963). Short-term memory in focal cerebral damage.

Pereira, A. E., Kelly, M. O., Lu, X., & Risko, E. F. (2022). On our susceptibility to external memory store manipulation: Examining the influence of perceived reliability and expected access to an external store. In *Memory Online* (pp. 44-60). Routledge.

Raaijmakers, J. G., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological review*, *88*(2), 93. https://doi.org/10.1037/0033-295X.88.2.93

Recht, S., Jovanovic, L., Mamassian, P., & Balsdon, T. (2022). Confidence at the limits of human nested cognition. *Neuroscience of consciousness*, *2022*(1), niac014. https://doi.org/10.1093/nc/niac014

Rhodes, M. G. (2016). Judgments of learning: Methods, data, and theory. *The Oxford handbook of metamemory*, *1*, 65-80.

Richardson-Klavehn, A., & Bjork, R. A. (1988). Measures of memory. *Annual review of psychology*, *39*(1), 475-543.

Risko, E. F., & Dunn, T. L. (2015). Storing information in-the-world: Metacognition and cognitive offloading in a short-term memory task. *Consciousness and cognition*, 36, 61-74. https://doi.org/10.1016/j.concog.2015.05.014

Risko, E. F., & Gilbert, S. J. (2016). Cognitive offloading. *Trends in cognitive sciences*, *20*(9), 676-688. https://doi.org/10.1016/j.tics.2016.07.002

Sachdeva, C. (2023). *Causes and consequences of cognitive offloading* (Doctoral dissertation, UCL (University College London)).

Sanders, E. C., & Berry, J. M. (2021). Evidence for an age-related positivity effect in metacognitive judgments. *The Journals of Gerontology: Series B*, *76*(7), 1282-1291. https://doi.org/10.1093/geronb/gbaa177

Scott, A. E., & Gilbert, S. J. (2024). Metacognition guides intention offloading and fulfillment of real-world plans. *Journal of Experimental Psychology: Applied*. https://doi.org/10.1037/xap0000515

Senkova, O., & Otani, H. (2021). Making judgments of learning enhances memory by inducing item-specific processing. *Memory & Cognition*, *49*, 955-967. https://doi.org/10.3758/s13421-020-01133-2

Serra, M. J., Dunlosky, J., & Hertzog, C. (2008). Do older adults show less confidence in their monitoring of learning?. *Experimental aging research*, *34*(4), 379-391. https://doi.org/10.1080/03610730802271898

Son, L. K., & Metcalfe, J. (2000). Metacognitive and control strategies in study-time allocation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(1), 204. https://doi.org/10.1037/0278-7393.26.1.204

Sperling, G. (1960). The information available in brief visual presentations. *Psychological monographs: General and applied*, *74*(11), 1. https://doi.org/10.1037/h0093759

Tauber, S. K., & Rhodes, M. G. (2012). Multiple bases for young and older adults' judgments of learning in multitrial learning. *Psychology and aging*, *27*(2), 474.

Thorndike, E. L. (1908). Memory for paired associates. *Psychological review*, *15*(2), 122. https://doi.org/10.1037/h0073570

Van Woudenberg, M. (2008). Optimal word pair learning in the short term: Using an activation based spacing model. *Unpublished master's thesis, University of Groningen*.

Yang, C., Sun, B., & Shanks, D. R. (2018). The anchoring effect in metamemory monitoring. *Memory & Cognition*, *46*, 384-397. https://doi.org/10.3758/s13421-017-0772-6

Young, J. (2012). *Personal knowledge capital: the inner and outer path of knowledge creation in a web world*. Elsevier.

Zheng, Y., Recht, S., & Rahnev, D. (2023). Common computations for metacognition and meta-metacognition. *Neuroscience of Consciousness*, *2023*(1), niad023. https://doi.org/10.1093/nc/niad023

Appendix 1. Fifty-three English paired words from South Florida Free Association norms for the Learning Session in Study 1, Study 2, and Study 3.

Number	Cue	Target
1*	ATTENTION	GRAB
2*	CITY	CROWD
3*	DIGIT	DOUBLE
4*	MOMENT	SOON
5*	SPACE	MOUNTAIN
6	ABUSE	FIGHT
7	AGE	PEOPLE
8	ARTIST	DRAWER
9	BEAR	DOG
10	REPENTANCE	ADMIT
11	CARROT	HORSE
12	CEILING	LIGHT
13	CHISEL	BLOCK
14	COMPULSION	FORCE
15	CULTURE	GROW
16	DECENCY	GENTLEMAN
17	DELIGHT	PLEASANT
18	DISGRACE	EVIL
19	EASE	MIND
20	EXPRESSION	THOUGHT
21	FACT	HARD
22	FAITH	SONG
23	FIELD	DAY
24	FLY	ANNOYING
25	GLASS	JUICE

26	IMPRESSION	MOOD
27	IRON	DRESS
28	JEALOUSY	FIGHT
29	KITCHEN	BEDROOM
30	LEAN	STRONG
31	LUXURY	ITEM
32	MANNER	APPEARANCE
33	MOOD	ANGER
34	NEWSPAPER	PAPER
35	OPINION	INDIVIDUAL
36	PISTOL	WEAPON
37	POPE	HAT
38	POWER	GOD
39	PREFERENCE	OPINION
40	REASON	KNOWLEDGE
41	RESISTANCE	HOLD
42	REST	QUIET
43	ROAD	CONSTRUCTION
44	SAILOR	CAP
45	SCHOOL	YARD
46	SHOP	STOP
47	STONE	BONE
48	STUDENT	PERSON
49	SUPPORT	SELF
50	THEME	BOOK
51	THRESHOLD	HOUSE
52	WAR	FEAR

53	WELFARE	HEALTH

The number with “*” indicates that the pair is for the Practice session; the remaining pairs are for the formal sessions.

Appendix 2. Arithmetic operations for the Retention session.

Number	Question	Answer
1	1+2=	3
2	1+3=	4
3	1+4=	5
4	1+5=	6
5	1+6=	7
6	1+7=	8
7	1+8=	9
8	1+9=	10
9	2+1=	3
10	2+3=	5
11	2+4=	6
12	2+5=	7
13	2+6=	8
14	2+7=	9
15	2+8=	10
16	2+9=	11
17	3+1=	4
18	3+2=	5
19	3+4=	7
20	3+5=	8
21	3+6=	9
22	3+7=	10
23	3+8=	11
24	3+9=	12
25	4+1=	5
26	4+2=	6

27	4+3=	7
28	4+5=	9
29	4+6=	10
30	4+7=	11
31	4+8=	12
32	4+9=	13
33	5+1=	6
34	5+2=	7
35	5+3=	8
36	5+4=	9
37	5+6=	11
38	5+7=	12
39	5+8=	13
40	5+9=	14
41	6+1=	7
42	6+2=	8
43	6+3=	9
44	6+4=	10
45	6+5=	11
46	6+7=	13
47	6+8=	14
48	6+9=	15
49	7+1=	8
50	7+2=	9
51	7+3=	10
52	7+4=	11
53	7+5=	12

54	7+6=	13
55	7+8=	15
56	7+9=	16
57	8+1=	9
58	8+2=	10
59	8+3=	11
60	8+4=	12
61	8+5=	13
62	8+6=	14
63	8+7=	15
64	8+9=	17
65	9+1=	10
66	9+2=	11
67	9+3=	12
68	9+4=	13
69	9+5=	14
70	9+6=	15
71	9+7=	16
72	9+8=	17
73	1*2=	2
74	1*3=	3
75	1*4=	4
76	1*5=	5
77	1*6=	6
78	1*7=	7
79	1*8=	8
80	1*9=	9

81	2*1=	2
82	2*3=	6
83	2*4=	8
84	2*5=	10
85	2*6=	12
86	2*7=	14
87	2*8=	16
88	2*9=	18
89	3*1=	3
90	3*2=	6
91	3*4=	12
92	3*5=	15
93	3*6=	18
94	3*7=	21
95	3*8=	24
96	3*9=	27
97	4*1=	4
98	4*2=	8
99	4*3=	12
100	4*5=	20
101	4*6=	24
102	4*7=	28
103	4*8=	32
104	4*9=	36
105	5*1=	5
106	5*2=	10
107	5*3=	15

108	5*4=	20
109	5*6=	30
110	5*7=	35
111	5*8=	40
112	5*9=	45
113	6*1=	6
114	6*2=	12
115	6*3=	18
116	6*4=	24
117	6*5=	30
118	6*7=	42
119	6*8=	48
120	6*9=	54
121	7*1=	7
122	7*2=	14
123	7*3=	21
124	7*4=	28
125	7*5=	35
126	7*6=	42
127	7*8=	56
128	7*9=	63
129	8*1=	8
130	8*2=	16
131	8*3=	24
132	8*4=	32
133	8*5=	40
134	8*6=	48

135	8*7=	56
136	8*9=	72
137	9*1=	9
138	9*2=	18
139	9*3=	27
140	9*4=	36
141	9*5=	45
142	9*6=	54
143	9*7=	63
144	9*8=	72

Appendix 3. The participation consent.

"*Dear all participants,*

You are kindly invited to participate in this research.

This experiment has been approved by the Life Science Committee at Japan Advanced Institute of Science and Technology, a national institute established in 1990.

No harm will come to you from taking part in this experiment, but you may feel a mild fatigue by memorizing word pairs.

You have the right to stop at any time. Thank you for agreeing to take part in this experiment.

We are:

Mr. Yuan Ma, Doctoral Student, Email: yuan.ma@jaist.ac.jp

Prof. Dr. Tsutomu Fujinami, Supervisor, Email: fuji@jaist.ac.jp

We will be glad to answer any questions about this study at any time. The participant may contact us at the provided email address.

If you have questions about your rights in this research, or you have any other questions, concerns, suggestions, or complaints that you do not feel can be addressed by the researcher, please contact the person in charge at our institute:

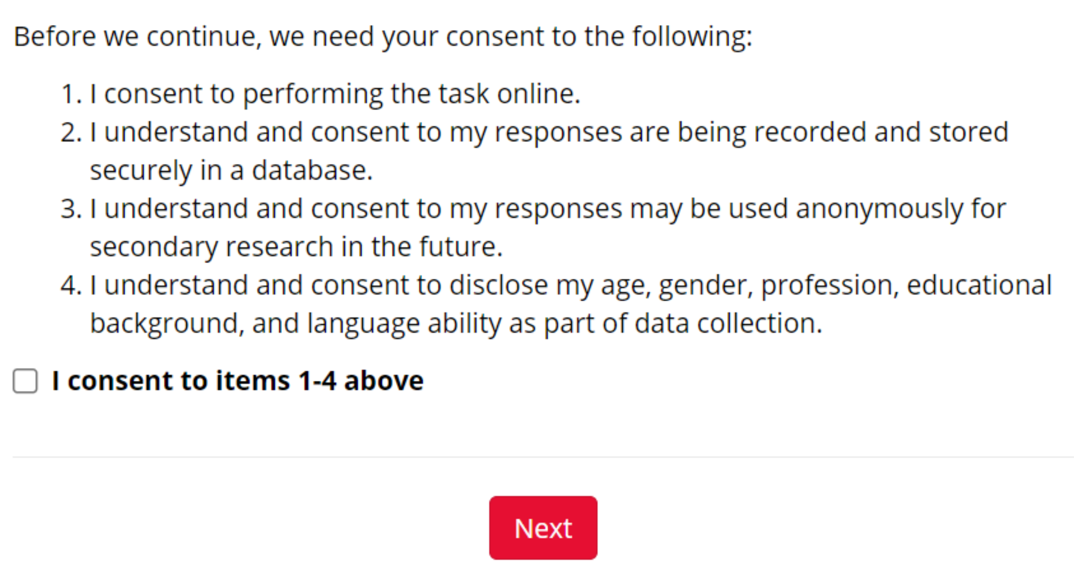
Prof. Asami Shikida (as-asami@jaist.ac.jp).

Before we continue, we need your consent to the following:

1. *I consent to performing the task online.*
2. *I understand and consent to my responses are being recorded and stored securely in a database.*
3. *I understand and consent to my responses may be used anonymously for secondary research in the future.*
4. *I understand and consent to disclose my age, gender, profession, educational background, and language ability as part of data collection.*

I consent to items 1-4 above"

Participants give this consent through Corllia on a webpage, see below:


Before we continue, we need your consent to the following:

1. I consent to performing the task online.
2. I understand and consent to my responses are being recorded and stored securely in a database.
3. I understand and consent to my responses may be used anonymously for secondary research in the future.
4. I understand and consent to disclose my age, gender, profession, educational background, and language ability as part of data collection.

I consent to items 1-4 above

Next

Appendix 4. Instructions for remote participation in Study 1 and Study 2.

First, participants read a general instruction:

"*Thanks again for your kindly agreeing to take part in this experiment.*

This project focuses on human memory performance in maintaining word pairs for later usage. The experiment consists of three sessions – Learning, Retention, and Testing.

In the Learning Session, you will learn 48 pairs of English words. Each pair is associated with two words in the upper case (e.g., ABILITY-CAPABILITY). Each pair is in one single trial. In each trial, you will see one pair.

The Retention Session provides a while to take a rest for your brain from memorizing those word pairs. You will finish simple arithmetical operations in about 10 mins. Some unexpected influence factors on your memory may be reduced (e. g., the order of stimuli) during this period.

Then is the Testing Session, in which you must perform as well as you can about recalling word pairs. Each item is one recalling item. You will see one word of the association as the cue in each item and input the other word of that association on your keyboard.

Additionally, please report your estimation about memorizing those word pairs in the Learning, Firming, and Testing sessions as a reference for the experimental result. Also, we need you to report your state after the experiment briefly. The full participation consumes about 50 mins. To become familiar with operating the program, you will have practice before the formal experiment.

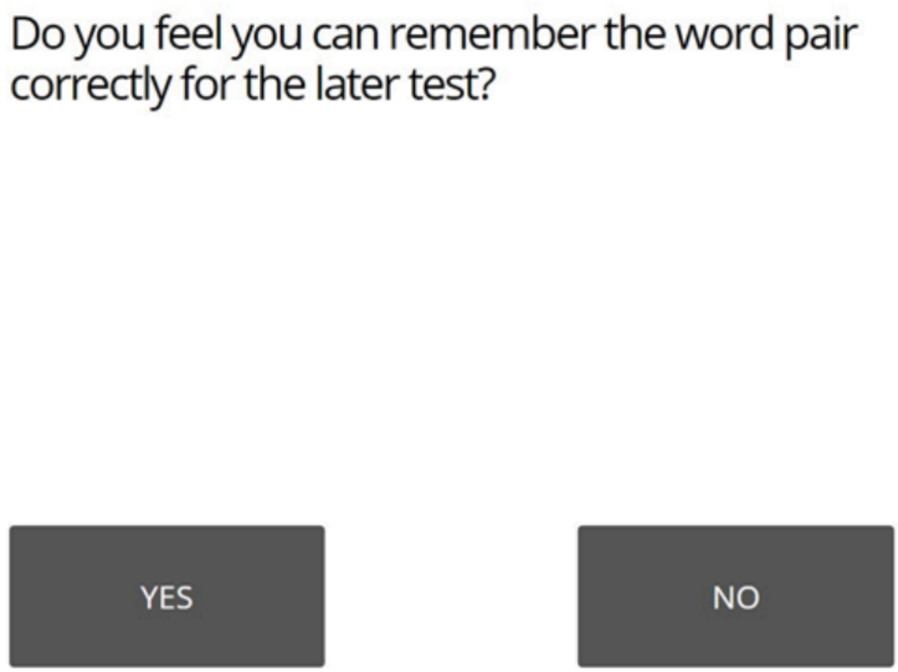
The detailed instructions with figures to show the entire experimental procedure are as follows. Please read it. Then you will start the Practice Session.
"

Then, participants view the detailed instructions on Gorllia, for Study1, shown below:

In the Learning Session, you will first see the word pair. You have 5s to try to memorize the association. It will automatically advance to the next screen then.

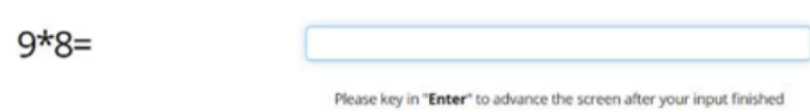
SPACE MOUNTAIN

Then, you will estimate your memory performance on learning that pair. Please submit your answer by clicking one of the two options ("YES" and "NO").

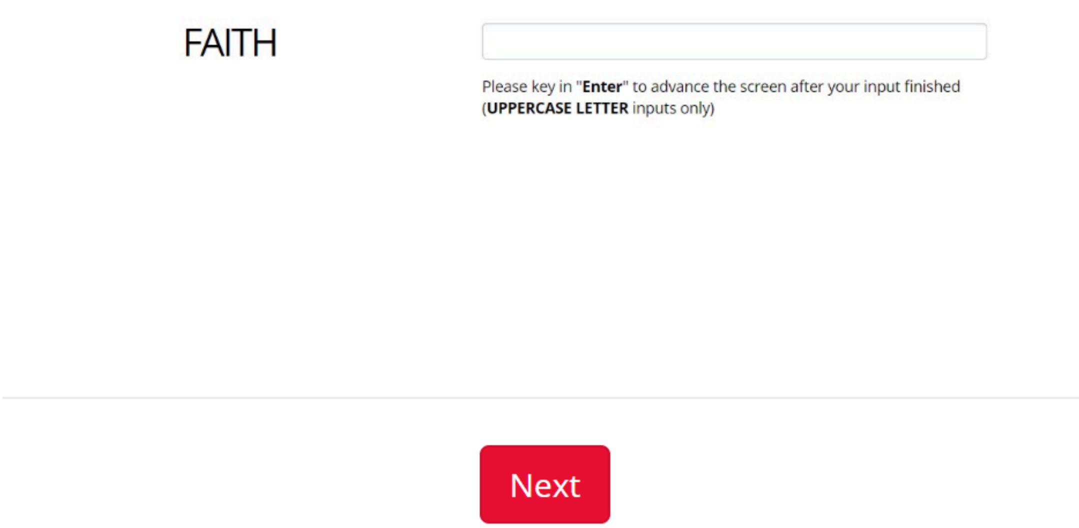


Next, you will report your confidence about your previous estimation on your memory performance. Please rate your confidence level by sliding the scale from 0 to 100 to indicate a particular value of your state, then click “Next” on the bottom middle.

In the Firming Session, in each trial, you need to figure the answer to each item on the left of that screen and input the value on the right entry area by a keyboard. Then, please input the “Enter” key in the entry area to advance the screen.



In the Testing Session, in each trial, you need to see the cue part of one word pair on the left side and provide the other part of that pair by inputting the answer into the entry area on the right side. There is no time limitation for your input, so please input the “Enter” key in the entry area to advance the screen. Please note that **only the CAPITAL inputs** can be recognized as correct answers.



The data we collect

We record **your responses** to the above test. We also ask you to inform us of your **age, gender, profession, educational background, and language ability** because these factors may influence how you remember things.

Please input your age below:

Please input your gender below:

Please input your profession below:

Please input your educational background below:

Please select your English language level:

Please Select...

Are you bilingual and multilingual?

Please Select...

Next

This program is a one-time test. Therefore, we ask you to open it for the first time. Please **confirm** below:

"This is the first time I have participated in this experiment."

Please Select...

Please note that dishonest confirmations may void the reward.

Next

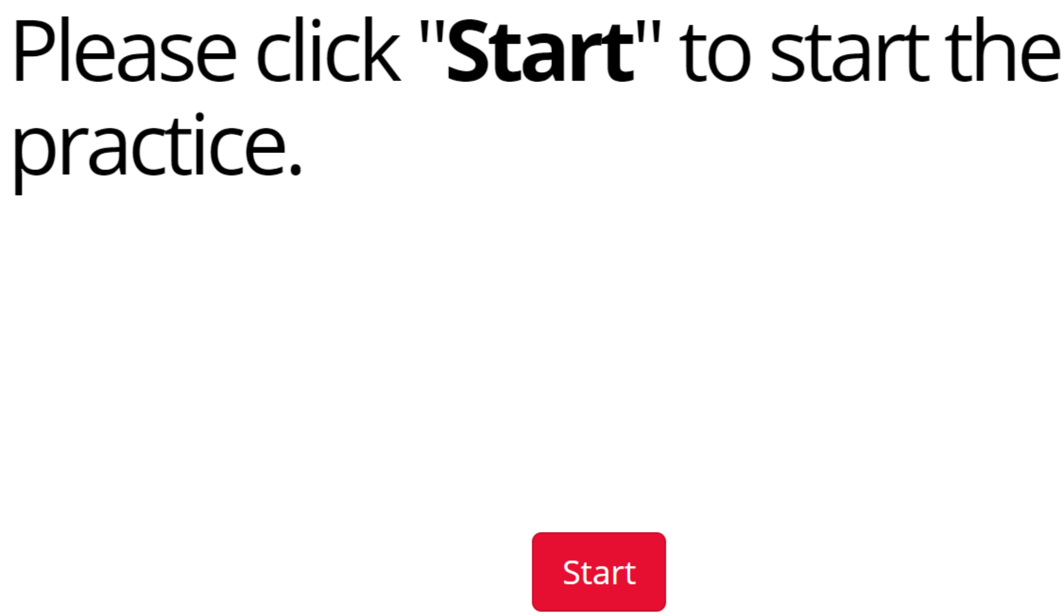
Also, we ask you to finish this test only by your memory, without using any additional support (e.g., writing memos, taking photos, etc.). Please **confirm** below:

"I will only use my memory ability to finish my participation."

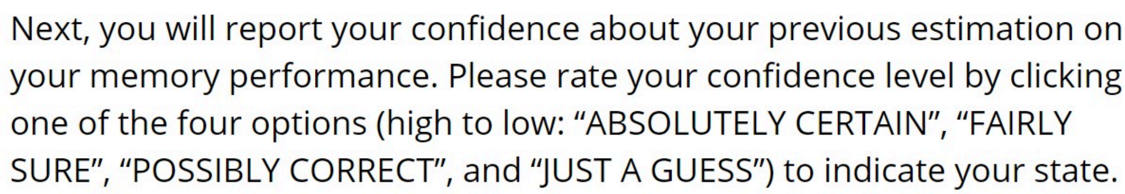
Please Select...

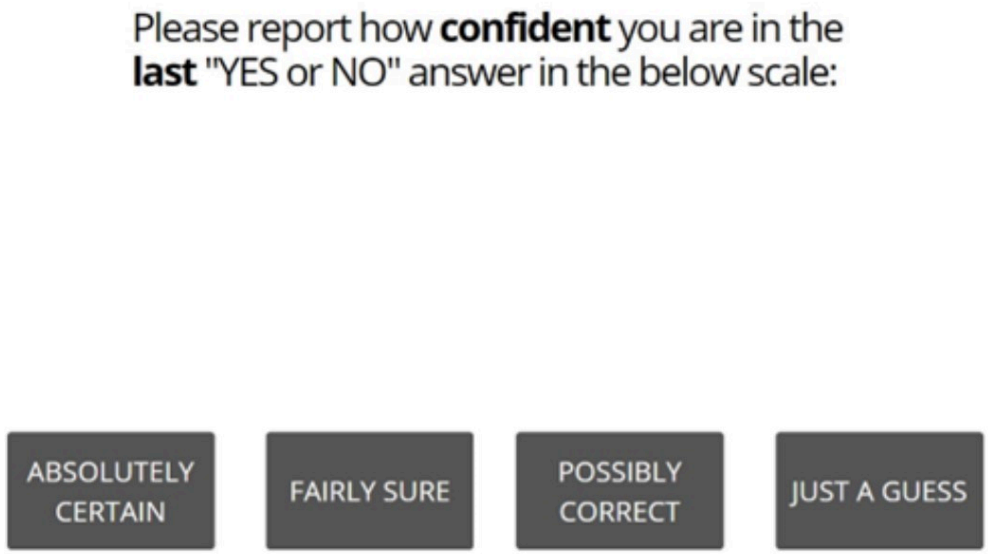
Please note that dishonest confirmations may void the reward.

Next



For Study 2, the difference is the screen used to submit second-order judgements, as shown below:





Appendix 5. Instructions for remote participation in Study 3.

First, the participants read a general instruction:
“*Dear all participants,*

Thanks again for your kindly agreeing to take part in this experiment.

This project focuses on strategic selections on maintaining word pairs for later usage. The experiment consists of three sessions – Learning, Retention, and Testing.

In the Learning Session, you will learn 48 pairs of English words. Each pair is associated with two words in the upper case (e.g., ABILITY-CAPABILITY). Each pair is in one single trial. In each trial, you will first see one pair. Next, you will choose to memorize the pair in your mind or write it down on paper. Finally, you need to report the reason for your choice.

The Retention Session provides a while to take a rest for your brain from memorizing those word pairs. You will finish simple arithmetical operations in about ten minutes. Some unexpected influence factors on your memory may be reduced (e. g., the order of stimuli) during this period.

Then is the Testing Session, in which you need to perform as well as you can about recalling word pairs. Each item is one recalling item. You will see one word of the association as the cue in each item and input the other word of that association on your keyboard. You will be allowed to refer to the paper if you write information down during the Learning Session.

Please be notified that we consider your choices in the Learning Session between two options ("Writing down" and "Keeping in mind") equally valid when we evaluate your performance in the Testing Session. We expect you to answer correctly as many as possible in the test.

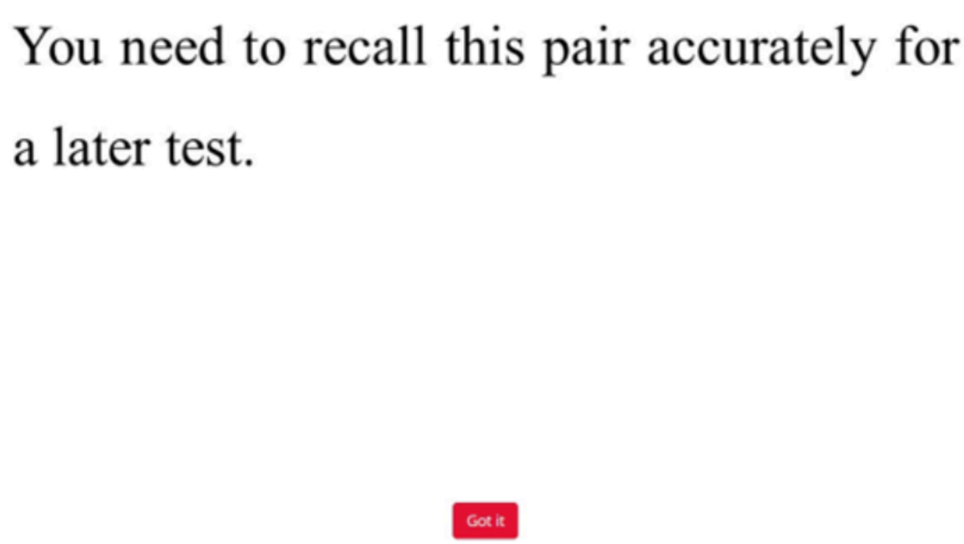
Additionally, we need you to report your estimation of your memory function in the Learning, Firming, and Testing sessions as a reference for the experimental result. The full participation consumes about 50 mins. Before starting the experimental program, please prepare a pen and paper near your computer to be available for writing word pairs down. To become familiar with operating the program, you will have practice before the formal experiment.

The detailed instructions with figures to show the entire experimental procedure are as follows. Please read it. Then you will start the Practice Session."

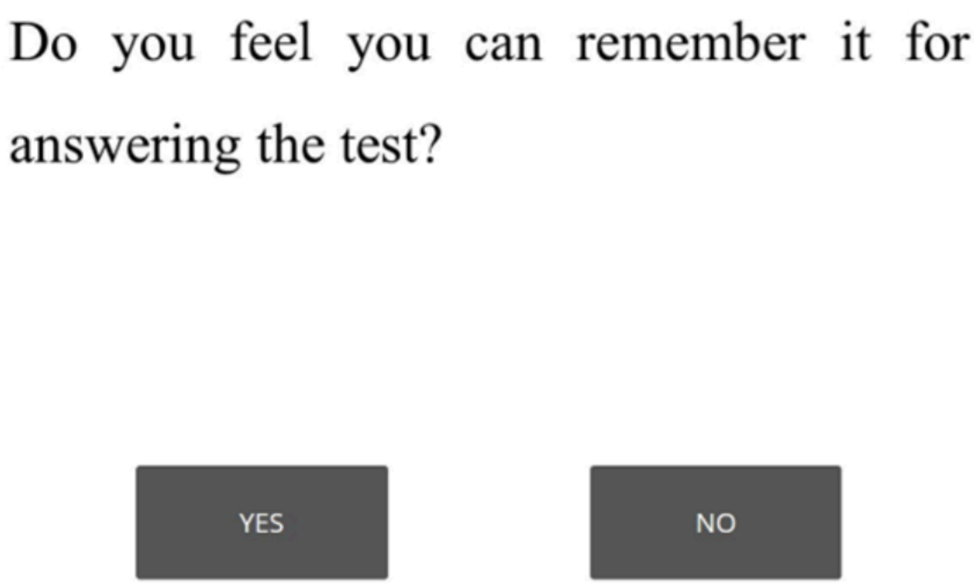
Then, participants view the detailed instructions on Gorilla, shown below:

In the Learning Session, you will first see the word pair. You have 5s to try to memorize the association. There is a countdown reminder on the bottom middle of the screen.

Then, you will see an instruction. Please click "Got it" on the bottom middle to advance to the next screen.

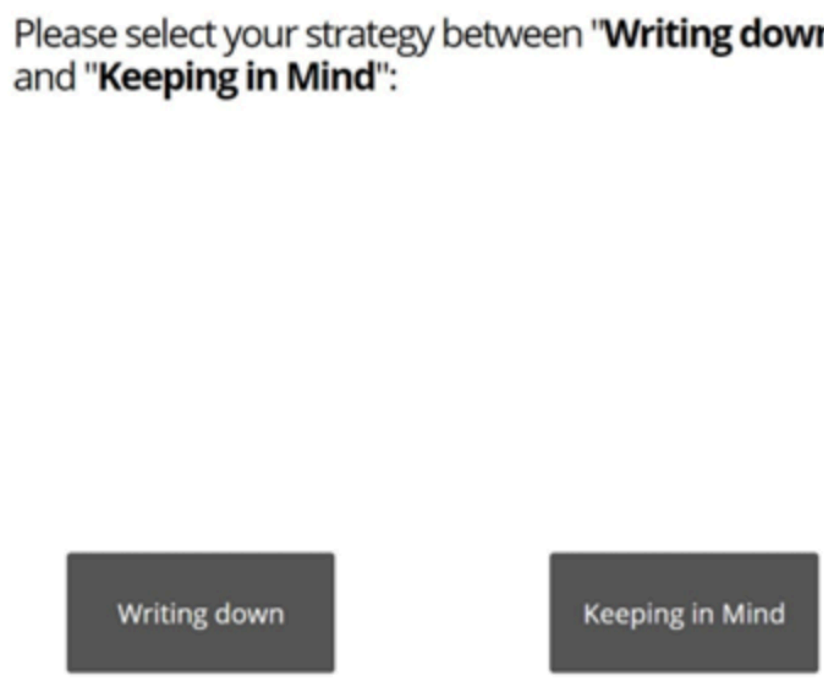


Next, you will estimate your memory performance on learning that pair. Please submit your answer by clicking one of the two options ("YES" and "NO").



Next, you will report your confidence about your previous estimation on your memory performance. Please rate your confidence level by sliding the scale from 0 to 100 to indicate a particular value of your state, then click "Next" on the bottom middle.

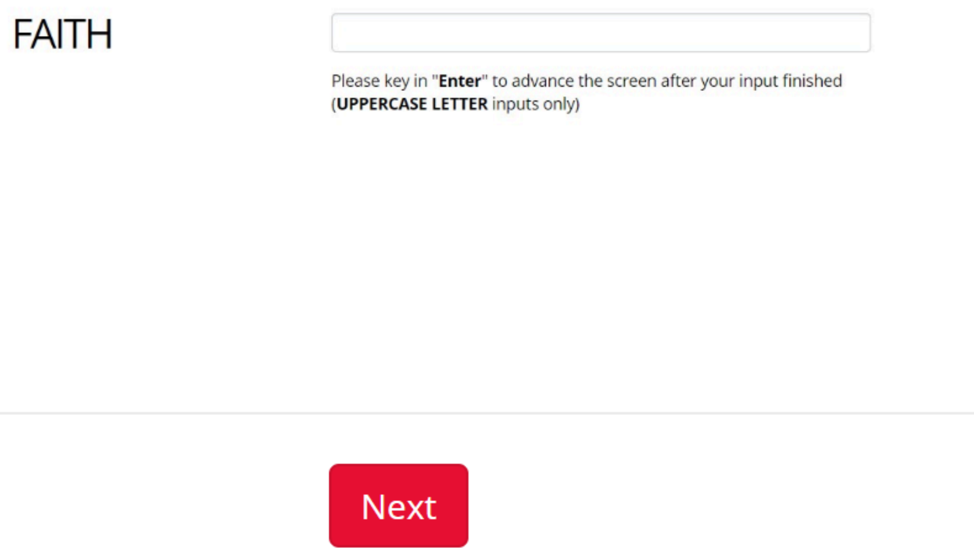
Next, you will make the strategic selection between memorizing the pair in your mind or writing it down on paper. Please submit your answer by clicking one of the two options (“Writing down” and “Keeping in Mind”).



If you select “Writing down” on the previous screen, you will see the word pair again. You can write the association down in this period. There is no time limitation, so please click “Written down” to advance the screen after finishing the record.

In the Firming Session, in each trial, you need to figure the answer to each item on the left of that screen and input the value on the right entry area by a keyboard. Then, please input the “Enter” key in the entry area to advance the screen.

In the Testing Session, in each trial, you need to see the cue part of one word pair on the left side and provide the other part of that pair (access memory or see the paper) by inputting the answer into the entry area on the right side. There is no time limitation for your input, so please input the "Enter" key in the entry area to advance the screen.



The data we collect

We record **your responses** to the above test. We also ask you to inform us of your **age, gender, profession, educational background, and language ability** because these factors may influence how you remember things.

Please input your age below:

Please input your gender below:

Please input your profession below:

Please input your educational background below:

Please select your English language level:

Please Select...

Are you bilingual and multilingual?

Please Select...

Next

This program is a one-time test. Therefore, we ask you to open it for the first time. Please **confirm** below:

"This is the first time I have participated in this experiment."

Please Select...

Please note that dishonest confirmations may void the reward.

Next

Also, we ask you to finish this test only by our provided options, namely your memory or writing on paper, without using other aids (e.g., digital devices, etc.). Please **confirm** below:

"I will follow this instruction during my participation."

Please Select...

Please note that dishonest confirmations may void the reward.

Next

Please click "**Start**" to start the practice.

Start